DOCUMENT RESUME

ED 295 573 IR 013 166

TITLE	Telecommunications Policy Research Conference. Standards and Standardization Section. Papers.
INSTITUTION	Telecommunications Policy Research Conference, Inc., Washington, DC.
PUB DATE	Sep 87
NOTE	136p.; Papers presented at the Telecommunications Policy Research Conference (15th, Airlie, VA, September 27-30, 1987). For the conference program and additional papers, see IR 013 160-182.
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Broadcast Industry; *Competition; *Design Requirements; Industry; Mathematical Models; Production Techniques; *Research and Development; Specifications; Standards; *Technological Advancement; Utilities
IDENTIFIERS	*Gateway Systems; *Standardization

ABSTRACT

These three papers consider various models and mechanisms for the achievement of industrial standardization. The first, "Duopoly Compatibility Standards with Partial Cooperation and Standards Leadership" (Sanford V. Berg, University of Florida) presents a model of duopolists producing differentiated substitutes, and considers the equilibrium outcomes for two products under Cournot rivalry, multiproduct monopoly, second-best standards-specification, and welfare maximization. The text is supplemented with two tables. The second paper, "The Economics of Gateway Technologies and Network Evolution: Lessons from Electricity Supply History" (Paul A. David and Julie Ann Bunn, Stanford University), examines the historical development of the technology of electric light and power systems, with particular attention to the role which the induced invention of a compatibility-creating technical device (the rotary converter) played in the industry's technological and organizational development. The text is extensively footnoted, and 71 references are provided. The third paper, "Coordination through Committees and Markets" (Joseph Farrell, University of California at Berkeley, and Garth Saloner, Massachusetts Institute of Technology), discusses two common mechanisms--standardization committees and "market leadership"--for achieving coordination, with particular reference to the choice of compatibility standards. The text is supplemented with figures, and is extensively footnoted. A list of 25 references is also provided. (Author/EW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED295573

TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE

Standards and Standardization Section

Airlie, Virginia

September 27-30, 1987

IR013166

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Edward Zajec
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Duopoly Compatibility Standards with Partial Cooperation and Standards Leadership*

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by Sanford V. Berg**

April 6, 1987
(Revised)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sanford V. Berg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

This paper presents a model of duopolists producing differentiated substitutes. Greater compatibility between the products causes each to have a greater demand. However, each firm has a most desired location in the product space (a most preferred technical standard): deviations from that standard in the direction of the rival's standard increases compatibility, but is costly. This paper presents the equilibrium outcomes for two products under Cournot rivalry, multiproduct monopoly, second-best standards-specification, and welfare maximization. The outcome associated with partial cooperation is examined. Then the Stackelberg model is used to derive the equilibrium conditions for output leadership and technical standards leadership. This model of technological externalities allows us to compare and evaluate different standards-specification processes.

*The author gratefully acknowledges comments received at the University of Florida Industrial Organization Workshop, especially from Leonard Cheng, Jonathan Hamilton, and Thomas Cooper. Comments on the framework by Roger Noll and an unknown referee were particularly helpful.

**Associate Professor of Economics, University of Florida and Executive Director, Public Utility Research Center.

IR013166

March 24, 1987

Duopoly Compatibility Standards with Partial Cooperation and Standards Leadership

by Sanford V. Berg

Several recent studies have emphasized how compatibility standards expand market demand by reducing uncertainty regarding the availability of related products (Hemenway, 1975; and Braunstein and White, 1985). For example, computers are part of complex systems which can include commercial software, storage devices, printers, and other components. The performance of computers, complementary goods, and associated inputs depends on the extent to which all components have technically compatible designs. This analysis concerns the decisions of firms that produce a basic product (like computers). Other firms -- not modeled here -- are expected to supply complementary goods competitively. In the case of computers, all else equal, consumers will prefer compatibility of operating systems for a number of reasons: to lower the costs of repairs (repairman can be less specialized and more will be competing for repair dollars), to have a larger selection of compatible complementary inputs (like software), and to reduce the risk of being stranded (left with a system that is no longer supported by the firm whose standards lose out).

The unique aspect of the theoretical model developed in this paper is that it characterizes the underlying technologies, including firm-specific features, and allows the interactions between rivalrous firms to generate an equilibrium degree of compatibility. An all-or-nothing outcome is avoided in this framework. Particularly in a market such as computers, many (but not necessarily all) complementary products designed for one computer can be used with varying degrees of success in conjunction with another computer, so we can have partial compatibility. Similarly, adaptors may be purchased which allow

one system to use components designed for the other; in such situations the demand for a technically incompatible computer is only dampened by the price (and inconvenience) of an adaptor. The key point is that the cost or quality of complementary products vary with the degree of compatibility, defined here as closeness of the technical standards.

The analytics of the model are similar to those in traditional models of product location. For example, in Economides' (1981) modification of Hotelling's model, duopolists select price and location -- as they compete with each other and an outside good. The existence of pure strategy equilibria depends on the valuation placed on the good relative to the per unit travel costs. If consumers in the middle of the line value the outside good more than their valuation of either product (net of travel costs), the markets will be separated and a noncooperative equilibrium can arise.

To incorporate an externality analogous to compatibility, consumer valuation for each product would also depend on the distance between the two firms (representing the degree of product compatibility). Thus, as one firm moved physically closer to the other, consumers of its product would value that product more. Furthermore, additional consumers would be obtained as the increased valuations outweighed travel costs. The valuations of consumers for the product produced by the other firm would also increase -- expanding its market as well. If equilibria exist under rivalry, the firms will tend to be too far apart relative to the efficient outcome (with this location externality). In its own location decision, neither firm takes the increased valuations of the other firm's customers into account. In addition, when one firm locates closer to the other, the market served by the other expands -- possibly touching and cutting into the mover's market. Such a development

(reflecting increased compatibility) increases rivalry between the firms, dampening incentives to locate close to each other.

The role of compatibility in this formulation is quite different from the role posited by Katz and Shapiro (1985). Their network externality characterization has consumer valuations dependent on the number of customers served by each duopolist -- not on the degree of incompatibility (reflected in the distance between the standards). In the network externality formulation, the source of the problem is that consumers do not take into account the impact of their consumption decision on the valuations of others. In the technological externality formulation, the crucial decisions are made by firms regarding their choice of technical standards. Their standards decisions do not take into account the impact on buyers of their rival's product. Since the latter characterization seems to capture some key features of the problem, it is useful to explore how such technological interdependencies can affect market performance.

Because of the highly stylized nature of the Hotelling location model, a slightly more general duopoly formulation will be developed in this paper. With compatibility viewed as a continuous variable, rivalry can lead to the under-provision of compatibility. Other market structures and behavioral assumptions lead to different outcomes. In the case of a multiproduct monopolist, for example, the optimal standards choices are made since the monopolist correctly evaluates the externalities involved in one product's location vis a vis the other. Nevertheless, its suboptimal output choice leads to overall performance which can be worse than that of the corresponding duopoly. Also, in the case of duopolists who cooperate only in the setting of technical standards, market performance may not be better than that achieved under complete rivalry. The technological externality perspective facilitates

the analysis of a number of interesting issues, including antitrust immunity for cooperation in the area of standards specification.

Here we present a model of compatibility standards that emerge in a duopoly. First we identify the welfare maximizing outcomes, and then present the second best solution -- when government can mandate standards, but cannot set output. After contrasting the results for rivalry and complete cooperation (monopoly), we analyze the outcomes under partial cooperation and leadership. As will be seen, the welfare rankings of the alternative standards specification arrangements depend on three factors: the impact on product cross-elasticities, the expansion of market demand, and the nature of the cost penalty for increased compatibility (as each firm deviates from its cost-minimizing production technique).

The framework provides some insights into the effects of cooperation and leadership on duopoly equilibrium. Partial cooperation corresponds to many real-world situations where firms collude to jointly determine the degree of compatibility for their products. While collusion that restricts output is a per se antitrust violation, cooperation in standards-setting is often condoned or even facilitated by government agencies, including the National Bureau of Standards. Another class of duopoly solutions emerges from the Stackleberg Leadership model, a reasonable characterization of the way in which standards evolve in industries with dominant firms. Even under symmetric demands and costs, one firm may be the acknowledged leader, and the other a technological follower. In such situations, the leader for the basic product incorporates the follower's reaction function into his profit maximization problem -- thus taking into account output (and standards) changes which would occur in response to the leaders' output and standards.

Although, the technological externality model captures very nicely some of the public policy issues associated with the selection of technical standards,

the analysis presented here does not examine strategic motivations (such as first-mover advantages), the impact of additional firms via further entry, or asymmetries in cost and demand. However, the model is versatile and incorporates the impact of increased compatibility on the substitutability between products and on consumers' marginal valuations (reflected in market demand). Thus, this framework for analyzing compatibility offers an alternative approach to the network model proposed by Katz and Shapiro (1985) and the bandwagon/inertia model developed by Farrell and Saloner (1985).

1. A Theory of Technological Externalities

For simplicity, let us assume a production function for each firm, with one traditional input (L_i) and a second (T_i) that reflects engineering protocols. The "technical standards" used by each firm have no effects on quality, but affect production costs. Costs are functions of outputs (y^i) and standards (T_i):

$$C^i = C^i(y^i;T_i) \; ; \; \partial C^i/\partial y^i > 0 \qquad (1)$$

For each y^i, there is assumed to be a unique T_i^* that minimizes C^i. Thus, if firm 1 adopts T_1^* and acquires L_1^* units of labor, costs will be minimized for the production of y_1^* units of output. Figure 1 illustrates the production isoquants: deviations from T_i^* raise costs:

$$\frac{\partial C^i}{\partial T_i} \quad \begin{matrix} < 0 \text{ for } T_i < T_i^* \\ = 0 \text{ for } T_i = T_i^* \\ > 0 \text{ for } T_i > T_i^* \end{matrix} \qquad (2)$$

We assume that for $y^1 = y^2$, $L_1^* = L_2^*$ but $T_1^* \neq T_2^*$ as shown in Figure 1. Without loss of generality, we assume that in equilibrium, firm 1 will always pick $T_1 \leq T_2$. Note that the formulation is quite general: T_i^* can be scale dependent

or independent of output level; marginal production costs can be dependent or independent of the deviation from T_i^*. As will be seen, separability of the cost function has implications for the degree of compatibility that arises under alternative market structures.

[Insert Figure 1 about here]

The inverse demand curve for either firm is a function of outputs and standards for both. Hence, revenue for firm i can be characterized as:

$$R^i(y^1,y^2;T_1,T_2) \tag{3}$$

Since the basic products are substitutes, these differentiated products may have special features that appeal to particular consumer-groups. However, greater output by rival firm j (and the corresponding lower price) affects firm i's demand. The quantity of y^i demand falls as y^j increases:

$$\partial R^i/\partial y^j < 0; \quad \partial^2 R^i/\partial y^i \partial y^j < 0 \tag{4}$$

Also marginal revenue, $\partial R^i/\partial y^i$, is greater or less than zero, depending on the output level. The nearness of T_1 and T_2 can be viewed as the degree of compatibility between these products. As the standards come close together, the demand for each product expands, increasing revenue:

$$\partial R^1/\partial T_1 > 0; \quad \partial R^2/\partial T_2 < 0; \quad T_1 < T_2 \tag{5}$$

1.1. Welfare Maximization

Consider a welfare maximizing multiproduct monopolist, where there are no economies of scope. Adopting a partial equilibrium approach, let z by the numeraire (priced at marginal cost) and assume constant marginal utility of income:

$$W(y^1,y^2,T_1,T_2) = z + U(y^1,y^2,T_1,T_2) - C^1(y^1,T_1) - C^2(y^2,T_2) \tag{6}$$

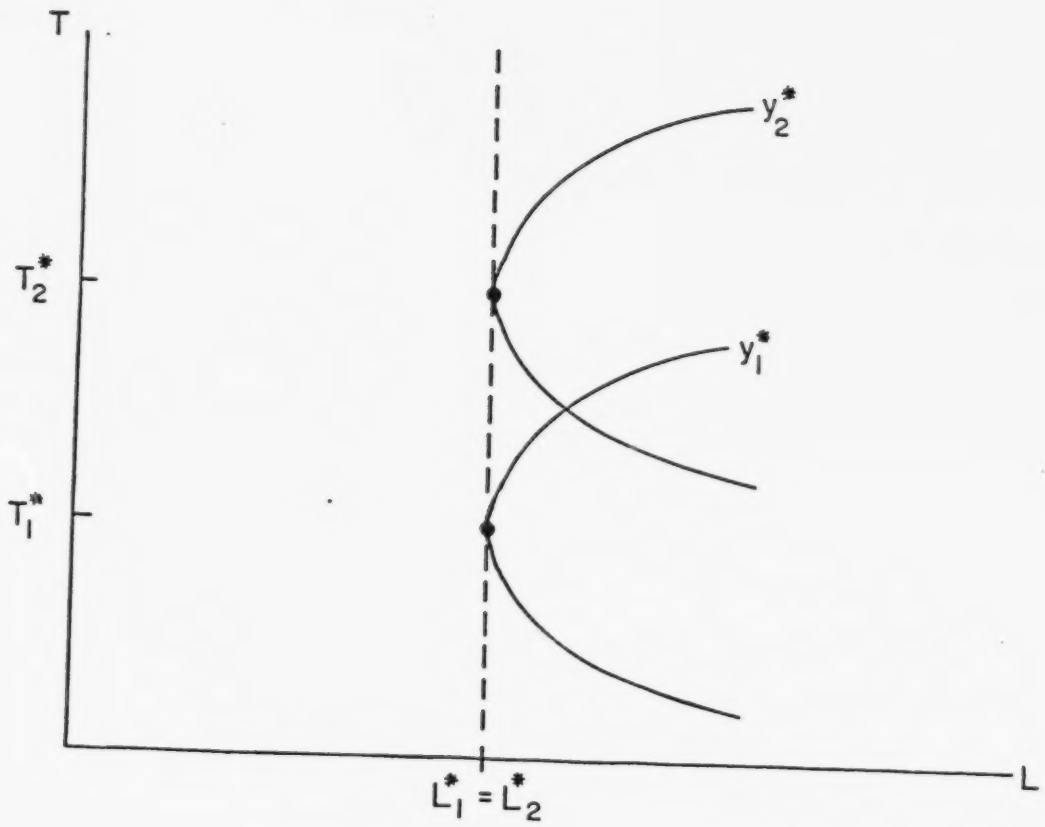

Figure 1

Cost Minimizing Standards under Symmetry Assumptions

The first order conditions involve equating the marginal benefits of additional output and changes in technical standards with the associated marginal costs:

$$\partial W/\partial y^i = \partial U/\partial y^i - \partial C^i/y^i = 0 \qquad (7)$$

$$\partial W/\partial T_i = \partial U/\partial T_i - \partial C^i/\partial T_i = 0 \qquad (8)$$

With four equations in four unknowns, one can solve for the optimal outputs and standards. The simultaneous solution for both technical standards allows the welfare maximizer to take into account the substitutability between products and the costs of deviations from the cost minimizing T_i^*. Note that the socially optimal output may involve producing only one product if product-specific fixed costs are very high, the goods are highly substitutable, and the demand penalty for incompatibility is great. For a two-product optimum, the degree of incompatibility (measured by $T_2 - T_1$) could be more than what emerges under monopoly or rivalry if the T_i^* diverge as output increases.

1.2 Second Best Standards Specification

If government can only set the T_i, then a second best solution results. The government agency would mandate T_1 and T_2, with output decisions subsequently made by the firms. If each firm also takes the rival's output as given, regulators would choose T_1 and T_2 to maximize welfare, subject to the profit maximizing output behavior of the firms. Standards-setting regulators are aware that each firm chooses y^i, which depend on T_i and T_j.

$$W = U(y^1(T_1,T_2),\ y^2(T_1,T_2),T_1,T_2) - C^1(y^1(T_1,T_2),T_1) - C^2(y^2(T_1,T_2),T_2) \qquad (9)$$

The FOC for standards are shown below:

$$\partial W/\partial T_i = (\partial U/\partial y^i)(\partial y^i/\partial T_i) + (\partial U/\partial y^j)(\partial y^j/\partial T_i) + \partial U/\partial T_i \qquad (10)$$
$$- (\partial C^i/\partial y^i)(\partial y^i/\partial T_i) - \partial C^i/\partial T_i - (\partial C^j/\partial y^j)(\partial y^j/\partial T_i) = 0$$
$$= (p_i - MC_i)(\partial y^i/\partial T_i) + (p_j - MC_j)(\partial y^j/\partial T_i) + \partial U/\partial T_i - \partial C^i/\partial T_i = 0$$

A Ramsey-like result obtains. Instead of equating the direct marginal benefits and marginal cost of, say, increasing T_1, the regulator sees through to the results of output rivalry. Thus, T_1 is increased up to the point where the additional cost of further deviating from T_1^* just equals direct benefits plus the additional benefits from narrowing the gap between price ($\partial U/\partial y^i = p_i$) and marginal cost ($\partial C^i/\partial y^i$) that occurs for each product as increases in T_1 induce output increases for both products.

To determine equilibrium output, we assume that firms maximize profits, where profits for each firm depend on outputs and standards of <u>both</u> firms:

$$\Pi^i(y^1,y^2;T_1,T_2) = R^i(y^1,y^2;T_1,T_2) - C^i(y^i,T_i) \qquad (11)$$

Assuming a Cournot output adjustment process, firms adjust output (and assume no reactions from the rival:

$$\partial \Pi^i/\partial y^i = \partial R^i/\partial y^i - \partial C/\partial y^i = 0$$

So again, we have four equations in four unknowns. The resulting system of equations allows us to determine equilibrium outputs and second-best standards when government only controls a one-time standards-making decision. Standards will tend to be closer together and output greater than under complete rivalry. Also, the second-best standards will tend to be closer than under unconstrained welfare maximization since they become an instrument for the promotion of output expansion. If greater compatibility increases the cross-elasticity of demand

between products, government can use standards-specification to promote rivalry. In addition, if the firms are going to collude on output, the government can partially offset the output restriction by bringing the standards even closer together. Note that setting $T_1=T_2$ to obtain complete compatibility will not, in general, be optimal unless the costs of deviating from T_i^* are very low and the marginal benefits extremely high.

1.3 Simultaneous Rivalry

If there is no government intervention, firms can be viewed as choosing T_i and y^i. In determining the equilibrium (y^i, T_i), we follow Brander and Spencer (1983) in their formulation of a two-step equilibrium for duopoly.[1] Elsewhere, the author (Berg, 1987) has presented the basic results for output and technical standards under rivalry so they will just be summarized here. Since the same approach is used to derive new results for partial cooperation and leadership, it is worthwhile to review the derivations.

Under rivalry, decisions are made independently; choices are based on available information and assumptions about rival responses. The Cournot assumption of unchanged y^j and T_j permits firm i to choose a profit maximizing combination of y^i and T_i. In choosing T_i, each firm must have some expectation regarding the ultimate resolution of the output rivalry. We assume that for any combination of T_i and T_j, firms correctly "see through" to the second step (output) equilibrium. The first step equilibrium is based on a Nash game in technical standards. The resulting "subgame perfect equilibrium" is such that all firms have their expectations confirmed.

---

[1] They analyzed R&D expenditures and output levels as the two choice variables. Note that changing from a Cournot output rivalry to Bertrand price rivalry would yield different results. So the conclusions are sensitive to this assumption.

We can derive the Cournot-Nash outputs and standards, under the assumption that firms simultaneously select T_i and y^i. Each firm assumes that the rival will adjust neither its technical standard nor its output. One first order condition for profit maximization in equilibrium is that marginal revenue equals marginal cost of production:

$$\partial \Pi^i / \partial y^i = \partial R^i / \partial y^i - \partial C^i / \partial y^i = 0 \qquad (12)$$

Also, an increase in T_1 expands the demand for both products, holding $T_2 (> T_1)$ constant. In equilibrium, the additional revenue from changing T must equal the additional cost:

$$\partial \Pi^i / \partial T_i = \partial R^i / \partial T_i - \partial C^i / \partial T_i = 0 \qquad (13)$$

Taking the first order conditions for both firms, we have four equations in four unknowns (y^1, y^2, T_1, T_2). This system can be solved for the equilibrium outputs and standards.

Interpreting the process is relatively straightforward. In the single variable Nash-Cournot model, output reaction functions determine the equilibrium outcome for rivalry. Similarly, for this two-variable model, we can envision reaction functions in the output and standards dimensions. With proper bounds on the revenue and cost functions, a unique, stable equilibrium will exist. However, without being more specific about functional forms and parameters (in contrast to the simple Nash-Cournot result), the existence of the technological externality (T_i enters into firm j's demand) means that a welfare ranking cannot be made between multiproduct monopoly and rivalry. In the case of standards rivalry, not only do firms not equate marginal benefits and marginal costs, but they do not take the impact of T_i on y^j into account. Table 1 displays the

first order conditions for welfare maximization, second best standards-specification, simultaneous rivalry, and monopoly.

[Insert Table 1 About Here]

Note that in the two-product equilibrium, the welfare maximizing compatibility standards will not necessarily be closer together than those developed under rivalry. The specific cost functions and demand interdependencies will determine the optimal T_i and T_j. The standards may depend on the scale of production, which (in turn) depends on marginal valuations of output. Using symmetric, quadratic utility functions, one can derive specific results for these four cases as well as those developed in the next sections. If the cost penalty for deviations from T_i^* becomes larger as the deviation increased, and marginal production costs are independent of T, comparisons among the various behavioral models are easier to make than if costs were nonseparable. This assumption is utilized in the Appendix to obtain illustrative results.

1.4 Multiproduct Monopoly

Turning to the case of monopoly, we assume that in equilibrium the firm elects to produce two products. The monopolist restricts outputs since it considers marginal revenue and product substitutability when determining outputs and standards for the two products; however, it takes into account the demand augmentation stemming from compatibility (as well as associated cost impacts). The welfare loss associated with monopoly depends on the relative strengths of these effects, and is larger when there are output scale economies. Also standards economies will be affected if increased compatibility allows input suppliers to achieve scale economies. Thus, the monopolist's recognition of demand augmentation from increased compatibility could cause closer standards (and greater market demand) than would arise under rivalry.

Table 1

First Order Conditions for Four Situations

Equilibrium Conditions

Output	Standards
Welfare Maximization	
$\partial W/\partial y^i = \partial U/\partial y^i - \partial C^i/\partial y^i = 0$	$\partial W/\partial T_i = \partial U/\partial T_i - \partial C^i/\partial T_i = 0$
Second Best Standards Specification	
$\partial \Pi^i/\partial y^i = \partial R^i/\partial y^i - \partial C^i/\partial y^i = 0$	$\partial W/\partial T_i = (\partial U/\partial y^i - \partial C^i/\partial y^i)\partial y^i/\partial T_i + (\partial U/\partial y^j - \partial C^j/\partial y^j)\partial y^j/\partial T_i + \partial U/\partial T_i - \partial C^i/\partial T_i = 0$
Simultaneous Rivalry	
$\partial \Pi^i/\partial y^i = \partial R^i/\partial y^i - \partial C^i/\partial y^i = 0$	$\partial \Pi^i/\partial T_i = \partial R^i/\partial T_i - \partial C^i/\partial T_i = 0$
Multiproduct Monopoly	
$\partial \Pi^M/\partial y^i = \partial R^i/\partial y^i + \partial R^j/\partial y^i - \partial C^i/\partial y^i = 0$	$\partial \Pi^M/\partial T_i = \partial R^i/\partial T_i + \partial R^j/\partial T_i - \partial C^i/\partial T_i = 0$

The net impact depends on the shape of the cost function in output and standards,[2] and the nature of the demand shift (particularly inframarginal vs. marginal impacts). The relationship between output and consumer valuations of compatibility affects the outcomes and degree of misallocation associated with the non-welfare maximizing models. In general, the conclusions derived from this framework are similar to those obtained in the analysis of market structure and product quality (surveyed by Schmalensee, 1979), although compatibility introduces an externality into the story.

2. Cooperation in Standards Specification

The possibility of voluntary cooperation in the establishment of T_i and T_j (with output rivalry) warrants our attention. If such technical cooperation shades over into jointly determined outputs, then we have the multiproduct monopoly outcome. So let us assume a sequence of decisions in which standards are jointly specified, then firms independently choose outputs.[3] Although the completely separate procedures used by firms for determining T_i and y^i are somewhat artificial, this stylized characterization does incorporate a fundamental difference between rivalrous and cooperative processes. Antitrust authorities and standards agencies, like the National Bureau of Standards, tend

---

[2] The term technological convergence (divergence) might be applied to situations where the cost minimizing T_1^* and T_2^* come closer together (further apart) as output expands. Such features of the underlying production function would influence the equilibrium standards for the cases reviewed here.

[3] There can be asymmetries in the benefits from cooperation if one firm bears the majority of the costs from deviating from cost-minimizing standards or one firm benefits disproportionately from demand augmentation. Such situations would require side payments to induce cooperation. With complete symmetry, the marginal benefits and costs are equalized for each firm -- reducing the pressure for side payments. Also, inframarginal net gains are the same.

to view joint determination of technical standards as beneficial to consumers. Cooperation at this stage can internalize the externality. The two firms could be viewed as turning over their standards decisions to a single agency which attempts to maximize joint profits -- recognizing that output rivalry will arise. Although firms are considering marginal revenues rather than marginal benefits, they might be expected to agree on a greater level of compatibility than would arise under complete rivalry. We examine below whether partial cooperation leads to improved industry performance relative to complete rivalry.

Before considering the impact of cooperation, we should note how sequential rivalry affects equilibrium outputs and standards (with the latter determining the associated degree of compatibility). Under sequential choice, firms independently choose standards, observe the outcome, and then choose output levels. A firm's demand (y^i) depends on its choice of standards, T_i, and the other firm's choice, T_j. The sequential choice of design standards allows firm i to take into account the impact of output expansion by firm j due to its choice of T_j (due to market demand augmentation). This model of rivalry has a different information structure than under simultaneous decision-making. The sequential choice model of rivalry results in the equilibrium standards tending to be further apart than in the case of rivalry under simultaneous decision-making. Firm i is aware that some of the private benefit from changing T_i is lost since the rival is now able to take T_i into account when determining its output. When we consider cooperation and leadership, such an information set is more realistic -- since the cooperating firms or the technological leader are in positions to take advantage of how rivals react to the standards.

Let us view each firm as choosing their T_i independently, then observing T_j. In the next stage of decision-making, the rivals independently choose y^i and y^j. All expectations are confirmed in the subgame perfect equilibrium.

Firm i's profit maximizing choice of T_i will depend on its expected output. So firm i can be viewed as solving for the y^i that emerges under rivalry. Decision-makers are viewed as "seeing-through" the first stage, and making their output decisions. The system is then solved by working backwards to identify the T's, which are dependent on each firm's demand and on cost penalties. The equilibrium outputs are dependent on the standards chosen. Thus, the profit-maximizing firm's profits can be characterized as

$$\pi^i = R^i(y^1(T_1,T_2),\ y^2(T_1,T_2),T_1,T_2) - C^i(y^i(T_1,T_2),T_i)$$

Profit maximization with respect to T_i, given T_j, yields a Nash equilibrium with FOC:

$$\partial\pi^i/\partial T_i = (\partial R^i/\partial y^i)(\partial y^i/\partial T_i) + (\partial R^i/\partial y^j)(\partial y^j/\partial T_i) + \partial R^i/\partial T_i$$
$$- (\partial C^i/\partial y^i)(\partial y^i/\partial T_i) - \partial C^i/\partial T_i = 0$$

Since $MR_i = MC_i$, this reduces to

$$\partial\pi^i/\partial T_i = (\partial R^i/\partial y^j)(\partial y^j/\partial T_i) + \partial R^i/\partial T_i - \partial C^i/\partial T_i = 0 \qquad (14)$$

Unlike under the information set for simultaneous rivalry, the firm's marginal revenue from changing T_i includes the indirect loss in its revenues due to firm j's expanded output as the standards come closer together. The FOC for output (shown in Table 2) is the same as under simultaneous rivalry. See Berg (1987) for conditions needed to ensure existence and uniqueness of an equilibrium.

In the case of partial cooperation, we also have sequential decision-making. The two firms meet and set standards so as to maximize joint profits, seeing through to the result of output rivalry. Alternatively, they could assign the task of standards coordination to an external agency, such as

an industry task force operating under the auspices of a standards specifiction institute. Joint profits for firms 1 and 2 can be expressed as

$$\begin{aligned}\pi^M = \pi^1 + \pi^2 &= R^1(y^1(T_1,T_2),y^2(T_1,T_2),T_1,T_2) \\ &+ R^2(y^1(T_1,T_2),y^2(T_1,T_2),T_1,T_2) \\ &- C^1(y^1,(T_1,T_2),T_1) - C^2(y^2(T_1,T_2),T_2)\end{aligned} \tag{15}$$

The FOC for standards for firm i can be expressed as the following:

$$\begin{aligned}\partial\pi^M/\partial T_i &= (\partial R^i/\partial y^i)(\partial y^i/\partial T_i) + (\partial R^i/\partial y^j)(\partial y^j/\partial T_i) + \partial R^i/\partial T_i \\ &+ (\partial R^j/\partial y^i)(\partial y^i/\partial T_i) + (\partial R^j/\partial y^j)(\partial y^j/\partial T_i) + \partial R^j/\partial T_i \\ &- (\partial C^i/\partial y^i)(\partial y^i/\partial T_i) - \partial C^i/\partial T_i \\ &- (\partial C^j/\partial y^j)(\partial y^j/\partial T_i) = 0\end{aligned} \tag{16}$$

Since $MR_i = MC_i$ and $MR_j = MC_j$ in equilibrium, we can rearrange terms and reduce the equation to

$$\begin{aligned}\partial\pi^M/\partial T_i &= [\partial R^i/\partial T_i - \partial C^i/\partial T_i] \\ &+ [(\partial R^i/\partial y^j)^{(-)}(\partial y^j/\partial T_i)^{(+)} + (\partial R^j/\partial y^i)^{(-)}(\partial y^i/\partial T_i)^{(+)} + \partial R^j/\partial T_i^{(+)}] = 0\end{aligned} \tag{17}$$

Assume, without loss of generality that $T_i < T_j$, so $i = 1$. If there had been simultaneous rivalry, the first bracketed term would have been equal to zero in equilibrium. That is, firm 1 would increase T_1 up to the point where the direct marginal revenue equals marginal cost. Even under sequential rivalry, only the first component of the second bracketed term would be taken into account. With partial cooperation, the joint profit-maximizing choice of T_1 must take into account the loss of firm 1's revenue due to firm 2's output expansion In addition R^j falls as y^i increases (in response to a change in T_i), dampening the incentive to bring T_i closer to T_j. Taking into account the direct increase in firm 2's revenue due to demand augmentation ($\partial R^2/\partial T_1$, an externality effect) may

when comparing equilibrium T_i under two regimes.

or may not offset these losses. Thus, ~~the second bracketed~~ another set of terms is taken into account in the cooperative standards-specification process: if the term is sufficiently negative, T_1 will not be increased as much as under sequential (or even simultaneous) rivalry. That is, recognition of the cross effects stemming from changes in T_i could lead the firms to reduce compatibility relative to the complete rivalry case under Cournot assumptions. If T_1 and T_2 are further apart, market demand will be less than under rivalry. Output of each firm would tend to be less than under complete rivalry, so welfare could actually be reduced by cooperation on standards. On the other hand, if direct effects on firm 2's revenues are large relative to the cross effects -- greater compatibility will be achieved by joint specification of T_1 and T_2. Of course, if a joint profit maximization approach to standards specification is to be feasible, a side payment scheme might have to be implemented.[4]

3. Stackleberg Leadership

The partial cooperation model corresponds to real-world situations where some collusion occurs. The collusion may be condoned (as is the case for compatibility standards) or be a per se antitrust violation (as with conspiracies to restrict output). Alternatively, leadership (rather than cooperation) may characterize a market. The Stackleberg Leadership model represents another important duopoly solution which is relevant for evaluating markets with dominant firms. Here we summarize how leadership affects the outcome; an example presented in the Appendix illustrates some key points of this model.

---

[4] A similar story can be told for case where duopolists are able to collude (cooperate) on output, but cannot (or will not) agree on standards, so the T's are chosen independently. This case is not as realistic as the previous one but it still illustrates how competition can increase or decrease welfare in the presence of partial collusion.

3.1 Output Leadership

First, let us review the results for a homogeneous product with constant marginal costs (c) and a linear demand ($p = A - b(y^1 + y^2)$). Rivalry yields total Nash-Cournot output ($y^1 + y^2$) of $(2/3)((A - c)/b)$ while the welfare-maximizing aggregate output is $(A - c)/b$. The case of Stackelberg leadership results in total output of $(3/4)((A - c)/b)$, one-third of which is supplied by the follower. Such an intermediate result (between rivalry and monopoly) does not necessarily obtain when there are more complicated cost functions and/or technological externalities (demand dependent on the rival's choice of standards).

Let firm 1 be designated as the output leader, so its profits equal

$$\pi^1 = R^1(y^1(T_1, T_2), h^2(y^1(T_1, T_2)); T_1, T_2) - C^1(y^1(T_1,T_2), T_1) \quad (19)$$

where $h^2(y^1)$ is the output reaction function for firm 2. Here, firm 1 does not assume that the follower holds output constant when it changes y^1. Sequential decision-making is a logical characterization of such situations. Firms independently choose T_1 and T_2, then outputs are set based on leadership. The process continues until equilibrium is achieved. The subgame perfect equilibrium concept yields a solution to the leadership problem.

The first order conditions for standards are the same as those from sequential rivalry (see Table 2). At the next stage, however, given T_1 and T_2, firm 1 picks y^1, where $y^2 = h^2(y^1)$. Then firm 2 picks the associated y^2. The resulting first order condition for firm 1 is

$$\partial\pi^1/\partial y^1 = \partial R^1/\partial y^1 + (\partial R^1/\partial h^2)\ (\partial h^2/\partial y^1) - \partial C^1/\partial y^1 = 0 \quad (20)$$

The terms are not the same as for the sequential rivalry case (shown in Table 1), where

$$\partial\Pi/\partial y^1 = \partial R^1/\partial y^1 - \partial C^1/\partial y^1 = 0.$$

For given standards, the Stackleberg leader will expand output more because it recognizes that firm 2 will adjust its output downwards.

The FOC for firm 1's standard is also shown in Table 2. Together with conditions for the follower, not shown, there are four equations in four unknowns. In general, the equilibrium standards are not the same for output leadership or total rivalry, although both involve standards rivalry. Rather than develop this version of the model further, let us turn to one which captures some of the issues of particular interest to those comparing various forms of rivalry with the outcome under government mandated standards.

[Insert Table 2 About Here]

3.2 Standards Leadership

In many situations, due to an invention or historical accident, one firm is acknowledged as the technical leader in an industry. In the Stackleberg Standards Leadership model, firm 1 takes advantage of its knowledge of the rival's reaction function to its standards. But now, since the closeness of standards directly affects the demand for both outputs, the formulation is much more complicated than under standards rivalry. The FOC for output are the same as found under sequential rivalry (with T_1 and T_2 given). Profit for firm 1 is expressed as

$$\begin{aligned}\Pi^1 = {} & R^1(y^1(T_1, S^2(T_1)), y^2(T_1, S^2(T_1)), T_1, S^2(T_1)) \\ & - C^1(y^1(T_1, S^2(T_1)), T_1) \text{ where } T_2 = S^2(T_1)\end{aligned} \tag{21}$$

Table 2

First Order Conditions for Sequential Situations

Equilibrium Conditions

Output	**Standards**
Sequential Rivalry	
$\partial\Pi^i/\partial y^i = \partial R^i/\partial y^i - \partial C^i/\partial y^i = 0$	$\partial\Pi^i/\partial T_i = (\partial R^i/\partial y^j)(\partial y^j/\partial T_i) + \partial R^i/\partial T_i - \partial C^i/\partial T_i = 0$
Cooperation in Standards	
$\partial\Pi^i/\partial y^i = \partial R^i/\partial y^i - \partial C^i/\partial y^i = 0$	$\partial\Pi^M/\partial T_i = [\partial R^i/\partial T_i - \partial C^i/\partial T_i] + [(\partial R^i/\partial y^j)(\partial y^j/\partial T_i) + (\partial R^j/\partial y^i)(\partial y^i/\partial T_i) + \partial R^j/\partial T_i] = 0$
Stackelberg Output Leadership (Firm 1 as Leader)	
$\partial\Pi^1/\partial y^1 = \partial R^1/\partial y^1 + (\partial R^1/\partial h^2)(\partial h^2/\partial y^1) - \partial C^1/\partial y^1 = 0$	$\partial\Pi^1/\partial T_1 = \partial R^1/\partial T_1 - \partial C^1/\partial T_1 = 0$
Stackelberg Standards Leadership (Firm 1 as Leader)	
$\partial\Pi^1/\partial y^1 = \partial R^1/\partial y^1 - \partial C^1/\partial y^1 = 0$	$\partial\Pi^1/\partial T_1 = [(\partial R^1/\partial y^2)(\partial y^2/\partial T_1) + (\partial R^1/\partial y^2)(\partial y^2/\partial S^2)(\partial S^2/\partial T_1)] + \partial R^1/\partial T_1 + (\partial R^1/\partial S^2)(\partial S^2/\partial T_1) - \partial C^1/\partial T_1 = 0$

Again, the reaction function reflects firm 2's cost conditions, which (in general) will depend on the output level.

The FOC for standards is complicated, even after combining terms and simplifying:

$$\partial\Pi^1/\partial T_1 = [(\partial R^1/\partial y^2)(\partial y^2/\partial T_1) + (\partial R^1/\partial y^2)(\partial y^2/\partial S^2)(\partial S^2/\partial T_1)] + \partial R^1/\partial T_1 + (\partial R^1/\partial S^2)(\partial S^2/\partial T_1) - \partial C^1/\partial T_1 = 0 \quad (22)$$

The Leader takes the Follower's reactions completely into account, including the indirect effect shown in brackets: the revenue impacts of firm 2's output responses to changes in the leader's standards.

Besides the direct effects of increased compatibility on net profits, the same as in the care of sequential rivalry, the leader takes into account the additional revenue from the follower's reaction to changes in T_1, $(\partial R^1/\partial S^2)(\partial S^2/\partial T_1)$. The Appendix derives the equilibrium standards for the standards leader to illustrate how the various cost and demand parameters affect the outcome. As would be expected, the leader tends to choose standards which are closer to T_1^* than under sequential rivalry. Market output could be greater than that obtained under rivalry, so this market structure could yield an improvement in performance relative to rivalry in the absence of standards leadership.[5]

---

[5] A parallel result for complete leadership in both output standards can also be derived using the multiple strategy model developed by Salop and Scheffman (1984). In their model, the dominant firm controls its price and a parameter (α) affecting its costs, the rival's costs and market demand. Complete leadership allows the dominant firm to take advantage of its knowledge of the rival's output and reaction functions. A limitation of their approach is the assumed homogeneity of output. Nevertheless, they show that control of two instruments may increase or decrease output. Aggregate welfare effects are also ambiguous: the dominant firm's profit-maximizing strategy may reduce output produced by a higher cost follower--with its increased producer surplus more than offsetting decreased consumer surplus. Impact on the rival's costs raises

The results for leadership of both output and standards are not shown in Table 2 since it is just a combination of the other two leadership models. The first order conditions incorporate full knowledge of both reaction functions. With the quadratic cost and demand functions used in the Appendix, the leaders' output will be greater than under sequential rivalry.

4. Conclusions

The standards-specification process raises a number of interesting (and important) problems. In the absence of cooperation (or collusion), rivalry will restrict the degree of compatibility to be less than optimal. In the case of sequential rivalry, each firm recognizes that some of the benefit from bringing standards closer together is lost since the other firm will subsequently take T_i into account when setting q^j. Thus, less demand augmentation is achieved than if the standards and output decisions were made simultaneously. Furthermore, fear of changes in cross-elasticities may result in rivals preferring less compatibility (to avoid the profit consequences of output rivalry).

The results raise questions about the welfare impacts of different behavioral patterns. In general, decision processes associated with the different models can increase or decrease welfare relative to rivalry, depending on the cost functions and the nature of the demand interdependencies (reflecting the underlying utility function). Thus, one cannot say for sure that cooperation on standards will leave consumers better off (relative to complete

---

(footnote 5 continued) the possibility that firm 1 would strategically choose $T_1 < T_1^*$ (which is outside the range we have considered). Even with such a predatory policy (choosing technical standards which are designed to raise a rival's costs), welfare could be greater than under rivalry with independent decision-making.

rivalry). This result is quite significant since antitrust policy seems to view groups that facilitate compatibility as relatively benign. Many analysts view organizations like the American National Standards Institute and industry associations as helping to eliminate inefficiencies stemming from incompatibility. However, the implications for welfare are not necessarily positive.

Given past presumptions under antitrust that standards cooperation is beneficial (so long as output collusion is avoided), the analytical framework presented here raises some important questions for policy-makers. It may be that Stackleberg Leadership or complete rivalry is to be preferred in some situations. Alternatively, government-mandated standards can be used in influence market performance. Although the duopoly characterization is highly stylized, and the undimensional nature of "standards" is simplistic, the model of technological externalities is a promising avenue for further analysis.

References

Yale M. Braunstein and Lawrence J. White, "Setting Technical Compatibility Standards: An Economic Analysis," Antitrust Bulletin, Summer 1985, 337-355.

Sanford V. Berg, "Technological Externalities and a Theory of Technical Compatibility Standards," revised Working Paper, University of Florida 1987.

J.A. Brander and B.J. Spencer, "Strategic Commitment with R&D: the Symmetric Case," Bell Journal of Economics, Spring 1983, 225-235.

Nicholas S. Economides, "The Principles of Minimum Differentiation Revisited," European Economic Review April 1984, 24(3), 345-68.

J. Farrell and G. Saloner, "Standardization, Compatibility, and Innovation" Rand Journal of Economics, Spring 1985, 70-83.

David Hemenway, Industry Voluntary Product Standards, Cambridge, Mass: Ballinger Publishing Company, 1975.

Michael L. Katz and Carl Shapiro, "Network Externalities, Competition, and Compatibility, American Economic Review, June 1985, pp. 424-440.

Steve C. Salop and David T. Scheffman, "Multimarket Strategies in a Dominant Firm Industry," Working Paper No. 100, Bureau of Economics, Federal Trade Commission, April 1984.

Richard Schmalensee, "Market Structure, Durability, and Quality: A Selected Survey," Economic Inquiry, April 1979.

Appendix

Since technological leadership may characterize a number of industries, it is instructive to investigate the outcome in some detail. We derive results using simple quadratic cost and utility functions, assuming complete symmetry:

$$U = Z + ay^1 + ay^2 - \tfrac{1}{2}(b(y^1)^2 + 2my^1y^2 + b(y^2)^2) - h(y^1 + y^2)(T_1 - T_2)^2 \qquad \text{(A-1)}$$

$$\frac{dU}{dy^1} = P_1 = [a - h(T_1 - T_2)^2] - by^1 - my^2;\ b > 0,\ 0 < m^2 < b^2 \qquad \text{(A-2)}$$

$$C^1 = cy^1 + f(T_1 - T_1^*)^2 \qquad \text{(A-3)}$$

From (A-1), as the standards approach each other, the price intercept increases and approaches a; h, the demand penalty parameter for incompatibility has a larger impact as T_1 and T_2 are far apart. As b increases, demand becomes steeper; m is the substitution parameter -- small m implies small cross-effects.

This formulation assumes that increased compatibility vertically shifts each firm's demand. For the same market output, market demand increases, but cross-elasticities are assumed unchanged. If increased compatibility causes the products to appear to be more homogenous, then m would approach b. The possibility that the demand for firm 1's product would be more vulnerable to changes in the rival's output is one major explanation why firms may choose to avoid demand-augmenting compatibility.

Equation (A-2) uses a separable cost function: both the cost penalty parameter, f, and T_i^* are independent of output; also marginal production costs

are constant. Furthermore, we let $T_1^* = \frac{1}{2}T_2^*$ to simplify the characterization of the reaction functions (see Figure 2). Before analyzing the results under leadership, it is useful to review the output and standards equilibria for rivalry. The output reaction functions (in y^1, y^2 space) are downward sloping, with equilibrium occurring where the output reaction functions cross. The positions of the curves depend on the equilibrium T_1 and T_2, where FOC require that $T_2^* - T_2 = T_1 - T_1^*$. Under complete cost and demand symmetry, compatibility is pushed equally by the two firms (in equilibrium). As can be seen in Figure 2, line AB represents the set of standards equidistant from each firm's preferred standards (T_1^* and T_2^*). The positions of the standards reaction functions (denoted $S^i(T_j;y^i)$) depend on output levels. In general, low cost penalties for deviating from T_j^* (reflected in small values of f) lead to an equilibrium closer to B. So standards rivalry still can result in a high degree of compatibility. Alternatively, if demand augmentation for closer standards is great (large values of h), the rivals also approach B ($T_1 = T_2$) in equilibrium.

The results for the Stackleberg situation are in line with our expectations. Standards Leadership allows firm 1 to take advantage of firm 2's reaction function-- so that in equilibrium, T_1 will be closer to T_1^*, reducing the leader's cost penalty and increasing profits for firm 1 (compared with sequential rivalry). Before solving for the leader's FOC, we need to derive firm 2's reaction function. We observe that

$$\partial\pi^2/\partial T_2 = 2h(T_1 - T_2)y^2 + 2f(T_2^* - T_2) = 0 \qquad \text{(A-4)}$$

$$T_2 = S^2(T_1) = \frac{hy^2}{hy^2 + f} T_1 + \frac{f}{hy^2 + f} T_2^* \qquad \text{(A-5)}$$

Thus, profits for the Leader can be expressed in terms of parameters and T_1 (given y^1 and y^2):

$$\pi^1 = \{[a - h(T_1 - (\frac{hy^2}{hy^2 + f} T_1 + \frac{f}{hy^2 + f} T_2^*))^2] - by^1 - my^2\} y^1 - cy^1 - f(T_1 - T_1^*)^2 \quad \text{(A-6)}$$

$$\partial\pi^1/\partial T_1 = -2h(T_1 - (\frac{hy^2}{hy^2+f} T_1 + \frac{f}{hy^2+f} T_2^*))(1 - \frac{hy^2}{hy^2+f}) y^1 - 2f(T_1 - T_1^*) = 0 \quad \text{(A-7)}$$

Therefore

$$T_1 = \frac{f}{f - h\phi^2 y^1} T_1^* + \frac{h^2 f \phi y^1}{(f + hy^2)(f - h\phi^2 y^1)} T_2^* \quad \text{(A-8)}$$

where $\phi = 1 - \frac{hy^2}{hy^2+f}$

As leader, firm 1 will tend to choose standards that are closer to T_1^* than otherwise--yielding an equilibrium south-east of AB in Figure 2. It achieves higher profits at firm 2's expense (since T_2 is closer to T_1^* than under rivalry). However, market output could be greater, so theoretically <u>both</u> firms could be better off under standards leadership than sequential rivalry, given the externality associated with standards.

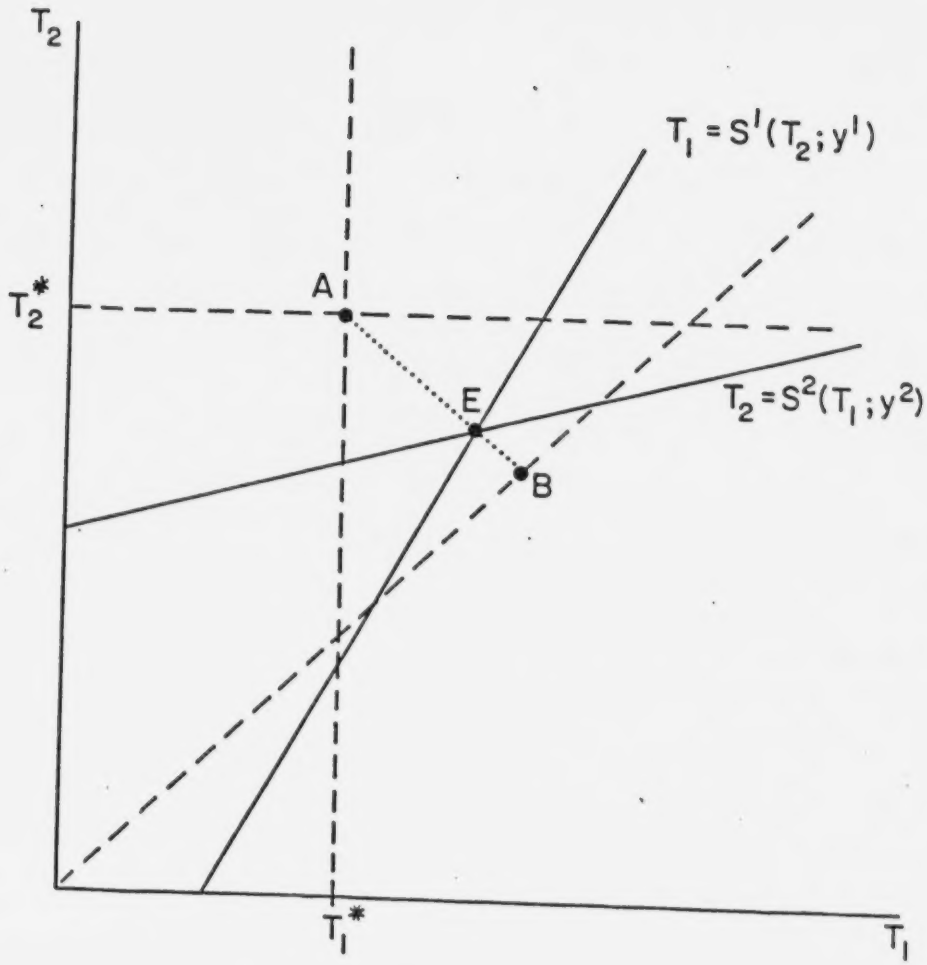

Figure 2

Equilibrium in Standards-Space
(Holding y^i Constant)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

THE ECONOMICS OF GATEWAY TECHNOLOGIES AND NETWORK EVOLUTION: LESSONS FROM ELECTRICITY SUPPLY HISTORY

by

Paul A. David and Julie Ann Bunn

Department of Economics
Stanford University

September 1987

This paper has been prepared for presentation at the Fifteenth Annual Telcommunications Policy Conference, at Airlie House, Airlie, Va., September 28-29, 1987

DRAFT: NOT FOR QUOTATION OR REPRODUCTION WITHOUT PERMISSION OF THE AUTHORS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

High Technology Impact Program

CENTER FOR ECONOMIC POLICY RESEARCH

100 Encina Commons
Stanford University
Stanford CA 94305

IR013166

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Paul A. David

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

Contemporary economic issues concerning technological innovation and compatibility standardization in emergent "network" industries are illuminated indirectly in this paper, by the examination of the historical development of the technology of electric light and power systems in the United States and Europe during the late nineteenth and early twentieth century. The discussion focuses on the rivalry between the initially incompatible technical formulations (a.c. vs. d.c.) of a system supplying electric lighting and power. In the dynamics of competition between alternative technologies under such conditions, the details of the timing of small historical events could have important and lasting consequences.

Special notice is given to the role which the induced invention of a compatibility-creating technical device (known to electrical engineers as the rotary converter) played in the conduct and resolution of the so-called "Battle of the Systems" between Edison and Westinghouse. The rotary converter is analyzed as a paradigmatic "gateway innovation", one of a large class of "minor" technological innovations whose actual importance in the context of network evolution warrants greater recognition from economists. Although technically neutral in its ability to convert either form of electric current into the other form, the rotary converter's effects upon the course of technological and organizational development in the industry are shown to have been surprisingly far-reaching in some respects, and not always unambiguously benign.

The Economics of Gateway Technologies and Network Evolution: Lessons from Electricity Supply History

by

Paul A. David and Julie Ann Bunn

0. Introduction

This essay explores some of the ways in which an appreciation of history may prove a useful guide for analysts of contemporary economic issues concerning innovation and standardization in the development of emergent "network technologies". By re-examining the dynamics of rivalries played out in the past among alternative formulations of the nascent electric-light and power supply technology, we may hope to illuminate the nature of the strategic choices that are likely to be faced today by technologists, entrepreneurs, and public authorities as a consequence of technical incompatibilities among components which might otherwise form more efficient, integrated systems of production.

The aspects of one paradigmatic network industry's history that will be related here are intended to be simply exemplary of many others that could be cited. They all would underscore the fact that increasing returns technologies which link their users in a "network", giving rise to positive externalities that increase with the extent of network use, typically involve coordinating the selection and operation of many complementary constituent parts. While some of these parts or sub-systems may be packaged together for use or purchase, it is rare when purely engineering considerations dictate that all the components of a complex network technology be supplied in a bundle by a single vendor.

Indeed, most of the large, multi-component and multi-agent systems of production with which we are familiar did not emerge full-blown in the forms that they, respectively, have come to assume. Only rarely will such systems be found to have fully and faithfully realized the initial integrated design concept of some single sponsoring agent. Rather, systems such as the railroads, electric-light and power utilities, and telephone networks should be thought of as being both society-shaping and "socially constructed".[1] These technologies have been built up sequentially, through an evolutionary process in which the design and operation of constituent components were adapted to the specific technical,

---

[1] See Hughes (1987) for this formulation.

economic and politico-legal circumstances in which new opportunities and problems were perceived. Those perceptions, usually, had been formed on the basis of experience acquired through the operation of pre-existing systems having some of the same functions, or ones directly analogous to the technology in question.

This points to the importance of a wide and varied array of factors--including elements so idosyncratic to one or another particular historical context as to appear from our perspective to have been pure matters of "chance"--in determining the precise timing of events in the developmental sequence. For, in the absence of perfect prevision on the part of the agents involved, it is the unfolding context formed by such events which conditions the selection of specific competitive strategies by the supplier of goods and services based upon rival technologies, and the decisions of potential and committed adopters of systems that lack universal compatibility.

Technological innovation itself, being in great part incremental and cumulative in its nature, is in each period conditioned by the prevailing context and, in turn, plays a part in shaping the economic decision-context that will come to supplant it in the sequence. Not only those options for innovative action that lie within the control of the sponsors of existing competitive technologies, but also the opportunities for entry by firms sponsoring new technical sub-systems which may be either complementary to, or substitutes for those already in the field, should be regarded in this light. Insofar as some perceived opportunities and problems are created by the existence of unintentional or designed technical incompatibilities which inhibit effective (and profitable) integration of potentially complementary sub-systems, the nature of ex post facto responses to the challenges thereby posed can exercise a critical influence over the subsequent course of network evolution.

The aspects of the development of electric-light and power technology during the late nineteenth and early twentieth century that we will be detailed here are of particular interest because they highlight a feature the evolution of large technological systems that appears to have escaped adequate recognition in recent microeconomic analyses of issues relating to compatability and standardization. This is the degree to which even the sponsors of alternative and incompatible technological systems may lack complete control over the course of the rivalry among them, owing to the possibility that other parties,

seeking to capture the benefits of network compatibility, may be induced to develop converter- or translator-innovations.

Innovations of this latter sort effectively open "gateways" permitting the technical interconnection of system-components that would otherwise remain isolated. Gateway innovations, however, generally will not be impartial in the influence they exert upon the evolution of a network industry. Indeed, as shall be seen, there are circumstances in which a converter whose technical functioning is essentially neutral--opening a "gateway" to passage in either of two directions--nevertheless may tip the balance of market competition decisively in favor of de facto standardization on one of the alternative technological systems rather than another. Such was the case of the "rotary converter", a seemingly innocuous device for transforming direct electrical current into alternating current and vice versa, whose reprecussions turned out to be far-reaching indeed.

Before proceeding with the narrative, some elements of the conceptual framework within which it should be be read need to be set out more explicitly. This is done in section 1, which discusses the key features of network technologies and some of their general economic implications, as well as the concept of technical "compatibility" and the functions of "gateway technologies." Section 2 then sets the stage for our historical drama by describing the emergence during the 1880's of the rival direct current and alternating current technologies of electricity generation, transmission and application. The advent of the rotary converter, as part of a cluster of innovations based upon the principles of ti e induction motor, is made the focus of the discussion in Section 3, which examines the role of this gateway innovation in the resolution of the so-called "Battle of the Systems" (in the U.S. market) in favor of a universal electricity supply system based on polyphase alternating current. Section 4 pursues a further and indirect reprecussion of the rotary converter, which contributed also to determining one of the technical parameters--cycle frequency--on which a.c. systems became standardized in the United States. Our conclusions are set out in Section 5, along with some qualifying observations on the extent to which the impact of technological innovations, including those which perform gateway functions, remains conditioned and constrained by the institutional and regulatory contexts in which the network industry is evolving.

1. Network Technologies, Gateway Innovations and History

The recent wave of improvements in information, computing and communications technologies has heightened popular awareness of some of the special opportunities and difficulties created by competitions waged by commercial sponsors of alternative technical formulations for new, complex products and production methods. Brief reference to our contemporary experience therefore can assist in forming an intuitive appreciation for the highly contingent way in which the technology of the electrical manufacturing and supply industries came to be elaborated at the end of the nineteenth century, and the significance of the rapid emergence in the United States of de facto technical standards for a universal electrical supply system based on alternating, rather than direct, current. Recall that during the early 1980s, the burning issue in the market for video tape recorders (VCRs, more specifically) was whether to commit to the Betamax format developed and sponsored by Sony, or cast one's lot with the VHS format group that had formed around the Japan Victor Corporation?[2]

Who among us has not wrestled with the question of whether our desktop computer is to be an IBM PC-compatible system, or an Apple system like the Macintosh? Or, paused to consider the call of the International Dvorak Society to reconfigure our computer keyboards, learn to touchtype on the DSK keyboard patented in 1932 by August Dvorak and W.L Dealey, and thereby escape a perpetual bondage to the inefficient QWERTY layout?[3] Such quandries seem all too commonplace nowadays.

Two features are significant about the prosaically modern choice-of-technique situations just instanced. First, they involve technologies characterized by increasing returns to scale of use or production, and second, they entail choices where considerations of technical interrelatedness among the

---

[2] A sequel now confronts the home video or "cam-corder" enthusiast, and the manufacturers of videocamera and VCR equipment: is the current half-inch (12.7mm) tape width for VCR's to be preserved as a standard, or is it better to switch to systems based on the 8 mm format pioneered by Sony? See Rosenbloom (1985) on the background; Cusumano (1985) on details of the VCR rivalry, from the producers' side.

[3] See David (1985, 1986a) for the story of QWERTY.

components forming alternative systems cannot be ignored. These are attributes of "network technologies", which, because they may give rise to pecuniary and technical externalities, create special problems as well as economic opportunities for private and public decision-makers. It is now quite widely recognized that users of a network technology are at the mercy of the social mechanisms available for maintaining efficient system performance by providing compatibility among all its constituent elements. Network technologies may, of course, be developed in a coordinated way through private commercial sponsorship of an integrated system: all the necessarily interrelated components of a product or production facility can, in many instances, be packaged together by a single provider who develops and maintains the required interface specifications. A difficulty can arise, however, when the entailed fixed capital costs are so large that they pose effective barriers to the entry of competitive formulations of the basic technology; the gains in operating efficiency assured by integrated supply then may be obtained at the expense of the inefficiencies that come with "natural monopoly" and too little diversity to suit the varied needs of end-users. On the other hand, when coordination of a network or technological system is left to the decentralized resource allocation processes of competitive markets, the latter have generally led to an insufficiently high degree of standardization or "compatibility" to avoid efficiency losses in systems operations.[4]

A different point, which has only lately come to be more fully appreciated among economic analysts, is that when private agents in pursuit of profits set about solving what seem to be reasonably straightforward problems of operating more efficiently with the technologies already at hand, the ensemble of suppliers and users can easily become committed to an unfavorable implicit "trade-off" against dynamic gains from further basic innovations in fundamental systems-

---

[4] The normative literature on the economics of standardization recognizes that while the evident failure of markets would call out for remedies in the form of governmental efforts to promote the exploitation of economies of scale obtainable via network coordination and system integration, such interventions may result all too easily in choices of technology that prove to be "mistakes." See Brock (1975), Kindleberger (1983), Carlton and Klamer (1983), Katz and Shapiro (1985a, b), Katz (1986), Farrell and Saloner (1985a, b, c), (1986). David (1986b) reviews this literature in greater detail.

design. When the relative attractiveness of adopting of a technical solution tends to increase as that particular practice or device gains adherents, and durable commitments are being made sequentially by different agents under conditions of imperfect information, the ultimate outcome of the process can turn on the detailed timing of small events--even happenings of a nature so extraneous as to appear random in character.[5]

Such commitment, moreover, is most likely to occur within a narrow frame of time, positioned near the beginning of the technology's trajectory of diffusion and development. What defines these critical phases, or windows in time, and causes them to close down quickly , essentially, is the growing weight that attaches to considerations of "network externalities"--the shaping of individual technological choices by the context formed through the prior decisions of other agents. For, it is the distribution of the already "installed base" that will come to count more and more heavily in determining choices about interrelated capital investments--which must be made in respect to both supplying and using the network technology.

Governmental intervention, therefore, is not the only potential source of "standardization errors"; markets also can make early and costly technological mistakes which subsequently prove difficult to undo. Under competitive conditions, and in the absence of public policy interventions, the existence of significant increasing returns to scale, or analogous positive feedback mechanisms such as "learning by doing," and by "using," can give the result that one particular formulation of a network technology--VHS-formatted VCRs, or QWERTY-formatted keyboards--will be able to drive out other variants and so emerge as the de facto standard for the industry. By no means need the commercial victor in this kind of systems rivalry be more efficient than the available alternatives. Nor need it be the one which adopters would have

---

[5] Under such conditions--where, as one of the present writers put it, more than a decade ago, "marked divergences between ultimate outcomes may flow from seemingly negligible differences in remote beginnings" (David (1975: p. 16)--market rivalries among variant technological systems do exhibit a strictly historical character, in the sense that they are non-ergodic; the dynamic process cannot shake itself loose from the grip of past events and its outcome therefore can properly be described as path-dependent.

chosen if a different sequence of actions on the part of others had preceded their own.[6]

Yet, by the same token it is entirely conceivable that for reasons having little to do with the ability of economic agents to foresee the future, or with anything more than an accidental and possibly quite transitory alignment of their private interests with the economic welfare of future generations, technological development may take a particular path which does prove eventually to have been superior to any available alternative.

This is a disquieting message for those who normally find comfort in the Panglossian presumption that "the invisible hand" of market competition--assisted sometimes by the visible hand of farsighted management--somehow has worked to shape ours into the most economically efficient of all possible worlds. Yet, recognition of the importance of network externalities naturally predisposes one to be extremely skeptical about the claims of that proposition to any general validity. We must acknowledge that where network technologies are involved one cannot justifiably suppose that the system which has evolved most fully is really superior to others whose development might have been carried further, but was not. Nor should we comfort ourselves with the presumption that the "right economic reasons" were responsible for the emergence a technological system that has in fact turned out to be superior to any of the alternatives available. These are two of the lessons borne out by the curious sequence of events to be recounted in the following sections of this paper.

There is another important, if rather less contentious message that will emerge from our narration of the early history of the electricity supply industry. It is simply this: rivalries between alternative and incompatible formulations of a new technology characterized by increasing returns to adoption, such as those which take the form of positive network externalities, may be resolved as a consequence of futher technological innovation. Although

---

[6] It should be emphasized that this form of (hindsight) recognition that rational decision-making had nonetheless resulted in an inferior path being mistaken for the "right" one, could not arise under conditions of constant or decreasing returns; in the latter cases the eventual outcomes would be path independent. See Arthur (1984, 1985), Arthur, Ermoliev, and Kaniovsky (1985) for more formal (mathematical) treatment, and David (1986b) for further exposition.

there are certainly enough commonplace examples to warrant the current fashion among microeconomic theorists of modeling the dynamic competition between technologies that will remain forever incompatible, it is time to add to the set of "stylized facts" one more very real possibility. The technical barriers separating the contestants may be overcome by an autonomous or induced innovation which introduces a "gateway technology", thereby making it feasible for suppliers or customers to integrate diverse sub-sytems into an enlarged production system or extended network.

What precisely is the meaning the term we have just introduced--**gateway technology**? Since the second quarter of the nineteenth century the neologism "gateway" has been used to indicate "a means of egress or ingress".[7] As employed in the present context the term retains the same general sense: **a gateway technology is some means (a device, or a coding convention) for effectuating whatever technical connections between distinct production sub-systems are required in order for them to be utilized in conjunction, within a larger integrated production system.**

A prosaic example of a gateway device is the adapter-plug which enables the modern Englishwoman to use her electric hair-drier (with its "three-point, square plug" fitting) on her holiday in Italy (where the sockets accept only plugs that are "two-point, round"). An equally plain illustration is the adapter-ring which allows one to attach camera lenses with bayonet-type mounts to a camera-body with a screw-type mounting. Still another is the nineteenth century contrivance of a railway car-wheel assembly featuring trucks with three (or more) wheels on a single axle, which allowed the same car to be used on railway lines of different gauges.[8]

---

[7] This comparative new-comer to the English language (which the O.E.D. on Historical Principles dates to 1842) has apparently proved useful in affording greater precision than "gate". The latter derived from Old English and Old Norse (geit, gat, gatan), and Teuton (gasse) words that, too ambiguously, denoted either a passage-way (as in Cannongate, a famous Edinburgh street), or to a moveable barrier obstructing an opening, which could serve as a means of regulating passage from or to some form of enclosure.

[8] See Puffert (1986) for this and other nineteenth century techniques developed to overcome the barriers created by differences in railway gauges.

All the forgoing are similar in that they overcome some pre-existing technical incompatibility which would otherwise prevent some particular set of components (camera-body and lens, electrical mains and hair-drying appliance) from being formed into an integrated production system. This is not the case with exemplars of pure "linkage" or communications technologies, such as the underwater cables linking the telegraph systems of different continents, or the modems and asynchronous communications support programs which enable computers to communicate via telephone lines with other computers or timesharing systems. It is a characteristic, on the other hand, of the contrivance (a connector cable and supporting software) marketed under the tradename "PC to MAC and Back", which is alledged to permit the transfer of data and text files between the IBM PC and the Apple Macintosh computer.

A gateway technology therefore may be said to be one which achieves **technical compatibility** in order to affect linkage or communication among subsystems. It will be noticed that, in the process of defining what is to be understood here by our use of the term gateway, the need has arisen to sharpen the concept of "compatibility" by indicating the economic content of that commonly employed term. Most recent economic analysis of "compatibility" standards problems considers technical compatibility to be a design property that permits physical devices "to work together", by either insuring that they fit together physically or electromagnetically, or enabling them to exchange reliably recognizable and reasonably unambiguous signals.[9]

---

[9] See, e.g., the survey by Farrell and Saloner (1986a): "We call products compatible when their design is coordinated in some way, enabling them to work together." This definition of compatiblity has strong ex ante connotations, reflecting the preoccupation--which most of the recent literature shares--with prospective standard-setting and design decisions made by technology suppliers. The term compatibility often is defined only implicitly, through examples, as in the reference by Katz (1986) to technologies that are "inherently incompatible (e.g., the color television technologies developed by RCA and CBS)". Besen and Johnson (1986:p.1) also avoid defining compatibility or incompatiblity explicitly, and instead send readers to Hemmenway (1975) for amplification of the distinction between standards relating to "compatibility" and other product attributes such as "quality". But usage of the term "compatibility" is more recent, and the word itself actually does not appear anywhere in Hemmenway's discussion (1975:Chs. 4, 5), which is focused upon the narrower conception of "uniformity standards for interchangeability". See David (1986b) for futher discussion of proposed taxonomies of standards and the relationships among them.

In fact, however, there seem to be two distinct contexts in which technical subsystems or components may be said to be compatible. In the first and more straightforward situation, technical components Y and Z are said to be compatible because they can be used with each other. Thus, the 35mm film and camera are "compatible"; the wordprocessing program Microsoft Word comes in two variants one of which is "compatible with" the Apple Macintosh computer, and the other of which is "compatible with" IBM PC's. But, ordinary language speakers are also heard to say that the IBM PC and the MacIntosh are "incompatible"--in being unable to read the same data-files or run the same programs. This latter is an instance of the second usage: component A and B are said to be "compatible" because each can be formed into a productive system with some third component, C. Here C could be a computer having DOS as its operating system, and A and B could be DOS-hosted applications programs; but, equally, A and B could represent two "PC-compatible" computers which ran the identical software, and could be directly linked in a local area network.

One way to avoid confusion is to distinguish between the two cases as being those of **compatible complements**, in the first instance, and of **compatible substitutes**, in the second. When there is a process for producing output with characteristics X which requires the integration of sub-systems (or sub-processes) A and B, the latter are compatible complements. System components A and B would be described as compatible substitutes (i.e., an IBM PC and a PC-compatible like Compaq which will run the same program) because there is a process for producing output with characteristics V, in which either (A,C) or (B,C) can serve as the inputs. It is worth noting that the same set of sub-systems or devices may, in different contexts, be described as compatible substitutes or complements; their relationship is defined by the technical nature of the production processes for X and V, respectively, which may have quite different output characteristics.

Further, as Berg (1985) has suggested, compatibility need not be conceptualized simply as being a "yes or no" proposition. Where the output of the production activity is multi-dimensional, components may be compatible complements (or substitutes) in the sense of being capable of cooperating with other components to yield output in one dimension, and yet be incompatible

with respect to performance in some other dimensions.[10] It may therefore be appropriate to regard compatibility as the achievement of conjoint performance which satisfies some minimal criteria (both as to levels and dimensionality), and to reserve the term "perfect compatibility" for the case in which two or more components functioning as compatible complements or (perfect) substitutes permit attainment of maximal system performance levels in all relevant output dimensions.

We are now in a position to state with greater precision what it is that gateway technologies do: **they make it technically feasible to utilize two or more components/sub-systems as compatible complements or compatible substitutes in an integrated system of production.** A "perfect gateway" device would permit the attainment of perfect compatibility, in the sense already defined. But, the availability of even a perfect gateway technology is not the same thing, from an economic standpoint, as the absence of any technical "incompatibility". There is no paradox intended here, just a recognition that to enable two otherwise incompatible sub-systems to function conjointly will impose some costs. These can take the form of fixed costs of installing a converter or translator, or they could appear as variable costs which depend upon how often, and with what speed or accuracy, the linkage or conversion operation is performed. Technical devices which are intended to function as gateways between systems that are in more-or-less continuous interaction, naturally, tend to be designed so as to minimize the marginal costs associated with conversion or translation, but, on that account, may impose significant fixed costs. Consequently, like other technological innovations, gateway innovations will tend to find their way into commercial application by private agents where the benefits derived from the incremental degree of system integration (and enhanced system performance) accrue to, and are in excess of, the costs they impose upon the agents who are in a position to adopt them.

---

[10] Not all computers ostensibly "compatible with" the IBM PC perform the identical range of operations, with identical results, when running programs written expressly for the PC. As a more mundane example, it turns out that the leader-strip on modern 35mm film rolls are cut at too sharp an angle to allow quick and dependable loading of some older vintage 35mm cameras (such as the Leica 3F), although once the leader has been taken up on the sprockets, all other operations such as winding and rewinding, and the positioning of numbered film frames, are performed perfectly.

The potential system-wide benefits of such innovations in the context of a developing network technology are obvious enough. Nevertheless, because those benefits may come mainly in the form of greater network externalities accruing to network users who have no direct need for the particular conversion capabilities that are offered, it is by no means necessary that market incentives will elicit installation of all the gateways that are technically feasible or socially optimal.

By extension of the latter argument, if network externality problems render the prospects for commercial adoption by private parties poor or especially uncertain, the profit incentives that otherwise might induce efforts towards the invention and development of gateway technologies, correspondingly, will be diminished. On the other hand, the direct advantages of achieving compatibility may well be suffiently concentrated in the hands of some identifiable group of potential network members that they constitute an attractive market --one that could profitably be supplied by the developer of a proprietary gateway technology.

The consequences of some new class of users being thus given access to existing networks can be far-reaching, not only in their immediate economic impact upon the costs and revenues of the sponsors of pre-existing incompatible systems, but also in terms of the direction imparted to the subsequent technological and organizational development of an enlarged network. Moreover, the economic sequlae of the creation of a new technological "gateway" are by no means unambiguously beneficial for the pre-existing class of users of the network facility. Some may suffer from elevated costs or reduced service as a result of network "congestion effects"; for others, whose specialized tastes or service-needs happen not to be shared by those admitted through the gateway, the economic tyranny of a new majority among network members may cause the future development of system capabilities to be deflected into channels which afford them few if any improvements of value.

Gateway technologies, and innovations therein, are sufficiently interesting and potent in the economic effects they may have upon the evolution of network industries that it is really rather surprising that to date so little notice has been given to them by economists concerned with theoretical, empirical and policy issues relating to network technology choices and compatibility standardization. This oversight may be explained in part by the fact that because the function of

a gateway technology is to overcome pre-existing incompatibilities, the analyst is required to envisage a specific, highly contingent technical context--with a particular structure of costs and prices--in which gateway innovations might appear. Very general theoretical conclusions are hard to derive from the study of such particular structures; the analytical exercise begins to look less like economic theory (as we have known it), and more like economic history. To explain an oversight, however, is not to excuse it. If the essential nature of the development of network industries is historical, as the foregoing discussion has suggested, we should try to come to grips with it.

2. The "Competition of the Currents"

The years 1887-1892 witnessed an episode of intense rivalry involving the proponents of two technologies vying for the electricity supply market. The contestants were, on one side, the incumbent direct or "continuous" current system sponsored by the complex of manufacturing and financial interests that had been built up around Edison's 1878 patent for a carbon filament incandescent lamp, and its sequelae;[11] and, on the other side, the challenging alternating current technology which at the time was represented on the American scene primarily by the Westinghouse and Thomson-Houston Companies. This was not a competition of long standing for, the first commercial a.c. system was built during the fall of 1886, in Buffalo , N.Y., by the Westinghouse Electric Co., an enterprise that had received its corporate charter from the State of Pennsylvania only as recently as January of the same year.[12] George Westinghouse, whose fame and fortune already had been established by more than a score of patents which he had secured for railroad air-brake apparatus in the years 1869-1873, was thus a comparative latecomer to the young electrical

---

[11] In 1882, according to Jones (1940), p. 41, this complex of enterprises included: (1) the Edison Electric Light Company, which had been formed to finance the invention, patenting, and development of Edison's electric-lighting system, and which licensed its use; (2) the Edison Electric Illuminating Company of New York, that being the first of the Edison municipal lighting utilities; (3) the Edison Machine Works, organized to manufacture the dynamos covered by Edison's patents; (4) the (Edison) Electric Tube Company, set up to manufacture the underground conductors for electric power distribution in the lighting system; and , (5) the Edison Lamp Works.

[12] See Prout (1921), pp. 94-5, 113; Sharlin (1963), pp. 195-6.

lighting industry.[13] In 1886 the Edison Electric Light Company already had been installing isolated plant electric supply systems for 5 years, and, having commercially implemented its system of incandescent lighting in 1882 by opening the Pearl Street Station, in New York City, and the Holborn Viaduct Station, in London, was well launched into the business of building "central" generating plants.[14]

Commercial exploitation of a.c. had awaited the demonstration--by the Frenchman, Lucian Gaulard, and his English business partner, J. D. Gibbs--that by using higher voltage alternating current and step-down transformers it was possible to achieve substantial reductions in the costs of moving energy, while making electrical power available at the low voltages then suitable for application to incandescent lamps.[15] The beauty of the thing was that transformers could be used essentially without loss of power to substitute voltage for amperage in order to reduce the need for the high conductivity, heavy gauge copper-wire transmission lines that were associated with d.c. lighting systems (thereby saving greatly on fixed capital costs), and then employed again in reversing the process to make low voltages available for local distribution at points of consumption.[16]

---

[13] On the pre-electrical career of Westinghouse, see Leupp (1919), Prout (1921).

[14] Fleming (1921), p. 225. Although Thomas Edison's inspiration had propelled direct current into an early position of leadership as the basis for a commercially implemented technology for electricity supply and application, the fundamental principles of generating alternating current with dynamos were not new. Indeed, the invention of the first direct current dynamo, by H. Pixii, in 1832, had involved the addition of a commutator to the alternating current dynamo he had built immediately following Michael Faraday's discovery of the "dynamo principle" in the preceding year. See, e.g., Sharlin (1963), pp. 136-47; Jarvis (1967a, b), for convenient reviews. More details of the history of alternating current technology prior to 1880 can be found in Electrical World (March, 1887); Sumpner (1890); Encyclopaedia Britannica (1910-11), vol. 9, p. 179-203; Fleming (1921), p. 44.

[15] Sharlin (1963), pp. 192-3; T. Hughes (1983), pp. 86-91.

[16] A "transformer," in modern electrical parlance, refers to a device by which the electromotive force (e.m.f.) of a source of alternating current may be increased or decreased. Unlike a dynamo used in generation, a transformer needs no moving parts. It has a primary winding (whose terminals are attached

There were, however, practical problems with the Gaulard-Gibbs transformer system, and it was not until 1885 that these were remedied by William Stanley, an American inventor in the private employ of George Westinghouse. The latter had become actively interested in the potentialities of alternating current during the previous year, and had moved quickly in securing the American rights to the Gaulard-Gibbs transformer patent. Within the same year, 1885, S. Z. de Ferranti was installing an a.c. lighting system with transformers wired in parallel in London's Grosvenor Gallery.[17] In the United States, commercial use of alternating current for lighting followed close on the heels of these developments. An experimental lighting system had been set up by Stanley in Great Barrington, Mass., and successfully demonstrated for Westinghouse in March, 1886; within six months Stanley's work was being put to use in Buffalo by the Westinghouse Electric Company.[18] Thus, on both sides of the Atlantic, Thomas Edison in 1886 was quite suddenly confronted by

---

to an a.c. source), and a secondary winding, both of which pass in coils around an iron "core." The first such device was the "closed-core" transformer made by Faraday, the core of which had the form of an iron ring. "Step-down" transformers have a larger number of turns of wire in the primary winding than in the secondary winding, since the ratio of primary (or "impressed") voltage to secondary voltage is equal to the ratio of the number of turns in the respective windings. Although it may seem , on first consideration, that the boost in voltage achieved by means of a step-up transformer somehow violates the conservation of energy law, such obviously cannot be the case; the power supplied at the primary is just equal to that delivered at the secondary. In general, when the voltage is stepped up (or, as English electrical engineers say, "bumped up"), the current is "bumped down" by the same proportion.

17 See Stanley (1912), pp. 564-65; Passer (1972), pp. 136-38, on Gaulard and Gibbs, Stanley and Westinghouse, and the still more comprehensive account in T. Hughes (1983), pp. 95-105. The challenge of solving the practical problems arising from the connection of the primary coils of the transformers in series by Gaulard and Gibbs, also drew a response in the same year from the Hungarian team of Zipernowsky, Deri, and Blathy of Budapest. Like Stanley, they found that transformers can be worked independently if the different primary circuits are arranged in parallel between two high voltage mains (1000v. in their patent), like the rungs on a ladder; the secondary circuits of the transformers were kept isolated, with incandescent lamps placed on them in parallel.

18 T. Hughes (1983), pp. 101-5; Passer (1972), pp. 131-8.

potential commercial rivals who had begun to explore an alternative technological trajectory.

Initially, the two systems did not compete directly. Instead, they staked out distinct portions of the market which were determined primarily by the differences in their inherent technical constraints. Two distance-related problems hampered immediate widespread application of the d.c. technology. First, despite implementation (beginning in 1883) of Edison's three-wire distribution design which reduced the cost of wiring by two-thirds relative to the previous two-wire scheme,[19] the cost of copper wire in the transmission lines continued to define what constituted an economically viable distance in competition with illuminating gas. Moreover, even if lower cost conductors were to be found, or new wiring systems devised, the low voltage d.c. system was distance-constrained on technical grounds having to do with "voltage drop". In other words, the further the distance a current travelled, the larger the voltage loss on the lines between the point of generation and the point of consumption. The earliest Edison systems, which typically were confined to a fairly small service area (approximately one mile in diameter), made allowance for this problem by generating at 105-110 volts in order to insure the delivery of at least 100 volts at the point of consumption--the latter being the voltage for which household lamps and other appliances were then being designed.[20] Beyond a certain transmission distance, however, this type of adjustment would no longer prove adequate; the system would deliver too broad a spectrum of voltages along the transmission lines.

Like others, George Westinghouse had seen these limitations of the direct current system as leaving an opening through which the alternative, Gaulard-Gibbs formulation could be used to make a significant inroad into the promising

---

19 Josephson (1959), pp. 231-2; Jarvis (1967b), Ch. 10, p. 229; T. Hughes (1983), pp. 83-4; Byatt (1979), p.99.

20 Jarvis (1976b), p. 229. The 110 volt standard for electricity consumption which came to be established in the United States was selected by Edison on the basis of electrical circuit theory and the maximum resistance lamp filament that he was able to obtain circa 1879. How the European's, following the Berliners' example, came to establish a 220 volt standard is another story. See Hughes (1983), p. 193.

electricity supply business dominated by Edison. Transmission of alternating current at higher voltages meant that with a given amount of generating capacity and a given weight of copper wire, the distance over which it remained economically feasible to deliver electric power would increase substantially. Indeed, for a given amount of power input, and a wire conductor of specified material and cross-sectional area, transmission distance increases as the square of voltage.[21] The economic implication of this was that expansion of supply from existing a.c. generating capacity--which allowed the further spreading of fixed capital charges--would be comparatively unconstrained by the costs of reaching customers over a wider area. Situating generating facilities at locations where ground rents were not as high as those found in urban central business districts, where electric energy demand for lighting, generally, was most concentrated, was another potential source of economic advantage. Transmitting alternating current at high voltage and then employing step-down transformers at sub-stations, or "bumping down" at the point of consumption, also greatly ameliorated the voltage drop problem encounter in operating d.c. systems.

Nevertheless, at this stage, Edison's system remained the better positioned of the two to dominate the more densely populated urban markets. Part of the reason was that there peak load lighting demands could be matched with Edison's larger, more efficient generators, leaving the a.c. supply systems to work the surrounding territory. Reinforcing this territorial division were numerous absolute disadvantages which blunted the ability of the early, single-phase a.c. technology to penetrate the urban market in competition with d.c.

At least four generic difficulties with a.c. remained to be resolved at the end of 1886. First was the fact--alluded to already--that the early Westinghouse and Thomson-Houston "alternators" (the a.c. equivalent to dynamo-generators) were only 70 percent efficient, whereas 90 percent efficiency was being achieved with d.c. dynamos, especially the large, "jumbo" design introduced by Edison. Second, the d.c. system was able to provide metered electric supply, whereas no a.c. meter had yet been developed--a deficiency that greatly reduced the system's attractiveness to central station operators. A third drawback for the a.c. technology at this stage of its development was that

---

21 See David (1987), p. 18 for further details.

although the principle of operating alternators in parallel had been demonstrated in 1883 (following the theoretical work of John Hopkinson), it remained to be translated into practical central station operations. The ability to connect dynamos in parallel rather than in series gave the d.c. system a distinct advantage: its generators could be disconnected and reconnected to the mains in response to varying load, thereby saving power inputs and wear and tear. Further, d.c. generators could be shut down for repair and maintenance--or might even be tolerated to break down--without disruption of the entire system.[22] Fourth, at a time when central stations employing the Edison system were beginning to spread fixed generating costs by supplying electricity for power as well as lighting, the a.c. system's ability to compete in urban markets was restricted by the lack of any satisfactory secondary motor available to be used with alternating current.

In addition to depriving electricity supply companies who chose to install alternators of the ability to serve industrial and commercial power customers, the lack of a motor became an increasingly pronounced comparative disadvantage with the success of the experimental electric streetcar systems installed during 1885 by Charles J. Van Depoele in New Orleans, South Bend, and Minneapolis. Its ability to deliver high torque at low r.p.m. made the direct current motor particularly well suited in such applications as traction, where continuous speed control was important.[23]

It should be emphasized that up until 1888 the underlying technological and economic considerations were not such as to allow great scope for the commercial realization of conventional scale economies and positive network externalities in the electricity supply business, as distinguished from the business of manufacturing electrical supply equipment and appliances, such as

---

22 Passer (1972), pp. 137-44,165-7; Byatt (1979), pp. 102-7; Evans (1892)p. 52.

23 The "traction boom" based upon d.c. was well and truly launched in 1887, with the completion of the superior streetcar system developed for the City of Richmond, Va., by Frank J. Sprague--a talented but erratic inventor formerly employed by Edison--in partnership with another Edison associate, Edward Johnson. There would be 154 electric street railway systems in operation in the United States by the close of 1889. See Passer (1972), pp. 237-49 on Sprague and Johnson particularly; pp. 216-55 on electric street railways; McDonald (1962), p.36.

lamps. Within local territories served by utility companies, of course, there were significant fixed cost requirements for generation and transmission, and these were sufficient to produce some "exclusion effects," or "first-mover advantages," which created incentive for racing between the sponsors of rival systems.[24] But, as has already been noted, rather than being symmetrically positioned with regard to every market, d.c. and a.c. systems each could find some markets in which they would enjoy some advantages in pre-empting entry by the other. Furthermore, such economies of scale as each could exploit in the generation of current remained definitely bounded by rising marginal distribution costs, with the geographical bounds due to transmission costs being more tightly constricted around the central generating stations in the case of d.c. systems. On the other side of the ledger, the possibilities of enjoying economies of scale in generation by achieving greater "load diversity", and consequently higher load factors for plants of given capacity, were far more limited for a.c. systems at this stage. The latter were still restricted to serving only the segment of the market (for incandescent lighting) that was characterized by very high peak-load in power usage. In the case of residential lighting demands, the load factor--defined as the ratio of the average load to the maximum load of a customer, group of customers, or an entire system during a specified period--typically might be as low as 10-20 per cent.[25]

Data on the number of central station and lamps associated with each of the rival formulations suggests that by the years 1888 and 1889 the diffusion of a.c. technology was rapidly catching up with that of the d.c. technology in the United States. As of October, 1888, the Westinghouse Electric Co. already could count 116 central stations (some of which, however, actually were d.c.

---

[24] And, also, between the main competitive suppliers of a.c. lighting plant-- the Thomson-Houston, and Westinghouse Electric Companies. On "first-mover advantages", and incentives for pre-emption by suppliers of new technologies and goods and services based upon them, see the theoretical analysis developed by Fudenberg and Tirole (1985); and Bresnahan and David (1986), for an effort at empirical application.

[25] See, e.g., T. Hughes (1983), pp. 217-19. A 10-12 per cent load factor is given as the average for a simple lighting load in the article contributed by J. A. Fleming on "Electricity Supply" in Encyclopedia Britannica (1910-1911), vol. 9, p. 194.

plants) with a total capacity to run 196,850 lamps, in comparison with the 185 Edison central stations operating 385,840 lamps. By 1891 a.c. had pulled ahead in the area of lighting, but, the available estimates of generating capacity for that date indicate that when electricity supply for power and traction is included, d.c. remained preponderant.[26] The catch-up that occurred during the latter 1880s mainly reflected the a.c. companies' move into the electrification of smaller cities and towns not well served by d.c., rather than penetration into the Edison system's "natural" territory in the spatially more dense urban lighting market.

There surely were some geographical markets in which the two variants were approximately balanced in their advantages, at least from the viewpoint of the cost of supplying electricity for lighting--larger territory for a.c. being offset by the impossibility of load balancing through connections to industrial power users, and traction companies. In such circumstances one well might expect that competition would feature the use of marketing tactics of all kinds, designed to tip the balance in one direction or another between competing system sponsors. And, in fact, it is precisely against this background of an apparent "technological stand-off" in the late 1880s that historians of technology and economic historians who have studied the electricity supply industry have set the ensuing episode known as the "Battle of the Systems."

As it has been recounted by more than one historian of technology and business enterprise, the "battle" that burst into public view during 1887-1892 was remarkable--even bizarre in some aspects--and regrettable in transgressing the normal boundaries for either market competition between different formulations of a new and highly promising technological paradigm, or professional disputation among scientists and engineers with regard to their comparative technical merits.[27] Spilling over from the market-place and the academy into the legal and political arenas, the "contest of the currents" between d.c. and a.c. took the form of courtroom struggles over patent rights, attempts to pass anti-competitive legislation, and public relations schemes aimed

---

[26] See Passer (1972), Table 19, p.150 for lighting capacity data for the three leading companies in 1891.

[27] See, for a sample of the best modern accounts in each genre, Stillwell (1934); Passer (1972), Ch. 5; T.Hughes (1983), Ch. 6, esp., pp. 106-111; J. Hughes (1986), pp.192-198. McDonald (1962), pp. 43-46 recounts the affair from the viewpoint of Samuel Insull.

at discrediting the opposition and frightening their customers. The conjuncture of these events in time has contributed to the impression that they all were facets of a concerted and unbridled counter-attack launched by the Edison camp, parts of an irrational effort to turn back the tide of technological progress that had brought an unwelcome influx of competition into the electricity supply business.

The patent litigation in these years was a largely coincidental aspect of the conflict,[28] and in any case not the one that has caused the "Battle of the Systems" to be characterized as bizarre. The most striking events of this episode revealed the lengths to which Edison and his immediate associates were prepared to go in order to convince the public that alternating current was an unsafe, basis for an electricity supply system. Their campaign was waged through a barrage of "scare" propaganda, supported by the grisly "evidence" Edison and his immediate associates produced by experimenting with a.c. at their laboratory in West Orange, N.J. As Edison's biographer, Matthew Josephson, relates:

> "There, on any day in 1887, one might have found Edison and his assistants occupied in certain cruel and lugubrious experiments: the electrocution of stray cats and dogs by means of high tension currents. In the presence of newspaper reporters and other invited guests, Edison and Batchelor would edge a little dog onto a sheet of tin to which were attached wires from an a-c generator supplying current at 1,000 volts."[29]

In July, 1888, Harold Brown, a former Edison laboratory assistant, electric pen salesman, and self-styled "Professor", put on a demonstration of the harmful effects of a.c. at Columbia College's School of Mines: the electrocution of a large dog was featured.[30] During 1888 the "West Orange gang," consisting of Edison, Johnson, and young Samuel Insull, aided by Brown and other assistants, succeeded in bringing off a related, stunning achievement in the art of negative promotion: they convinced the State of New York to substitute electrocution by administration of an alternating current for hanging, as the means of executing

---

28 See David (1987), pp. 23-26.

29 See Josephson (1959), p.347.

30 See Josephson (1959), pp.347-48; T. Hughes (1983),pp. 108-9; Woodbury (1960), p. 174.

convicted criminals. Edison himself had lobbied for this action before the New York legislature, and Brown--engaged as a consultant to the State on capital punishment by means of electrocution--sereptitiously purchased three Westinghouse alternators which he then announced had been selected as the type most suitable for such work.[31] The Edison group preceeded to milk the legislature's decision for all the publicity it was worth, circulating leaflets which, in warning the public about the dangers of high voltage a.c., used the term "to Westinghouse" in referring to lectrocution by alternating current.[32] Right through to 1889 Edison continued to participate in publicizing the undesirability of high voltage a.c. on grounds of safety, by criticizing the technical means given for rendering a.c. less hazardous.[33]

With the benefit of hindsight, and a seeming pre-disposition towards the "Whig" interpretation of technological history, previous chroniclers of this episode have been almost unanimous in portraying Edison as economically irrational in stubbornly championing the direct current system, and uncharacteristically, but nonetheless deplorably unscientific in his dogmatic public opposition to the rival alternating current technology sponsored by

---

31 George Westinghouse, reportedly, was outraged. One of these alternators was in place at the Auburn State Prison in August 1890 , where it produced the current lethal to the first victim of the "electric chair," William Kemmler, a convicted ax-murderer. See Leupp (1919), pp. 132-55 for details of the controversy over this execution, and the unfounded suspicions of the time that Westinghouse was financing the opponents of capital punishment who mobilized on the occasion.

32 See Josephson (1959), pp. 348-49; Cheney (1981), p. 43; T. Hughes (1976), p. 108.

33 In an article on "The Dangers of Electric Lighting," published originally in the North American Review, Edison stated that the undergrounding of high voltage line, (an often suggested means of increasing their safety) would serve only to make them more dangerous; further, he claimed that there was no means of effectively insulating overhead high-voltage wires, and concluded that "the only way in which safety can be secured is to restrict electrical pressure (voltage)." See Electrical World, November 2, 1889, pp. 292-3.

Thomson-Houston and the Westinghouse Electric Company.[34] Harold Passer described Edison's transformation bluntly in the following terms: "In 1879, Edison was a bold and courageous innovator. In 1889, he was a cautious and conservative defender of the status quo."[35] More severe is the judgement offered by Jonathan Hughes' spirited account, according to which, by 1886-1887, Edison

> had lost touch with the rapidly changing technology, or was fast losing touch. He had been a stanch, and then a rabid, opponent of alternating current transmission.... For some reason Edison could not comprehend the a.c. system. He was convinced that, transformers or not, the high-voltage of the a.c. system made it extraordinarily dangerous.... Arguments got nowhere with Edison on this issue. The fact that he considered the the main proponents of the a.c. system to be common thieves made him even more unwilling to see any virtue in their arguments.... There was no part of Edison's career that was so unworthy of the man and, in fact, sordid.[36]

Now, the reputation of Thomas Edison in American history-books is in all other respects so little in need of rehabilitation that there could scarcely be much call to repair it in this one regard, even if such a defense were plainly warranted by the facts. The concern here, accordingly, lies less with the way

---

34 See the influential account given by Josephson (1959), pp. 313-38, 349-350, 361-62, emphasizing Edison's resistance to the advice of others to follow Westinghouse and Thomson-Houston into the a.c. electricity supply field.

35 Passer (1972), p.74. This summation in Passer's 1954 study is quoted approvingly in the influential biography by Josephson (1959), p. 350. Alterations in the inventor's objective economic and institutional circumstances are cited by Passer in explanation of this transformation, whereas Josephson attributes Edison's unwillingness to pursue developments in a.c. technology to "the fear that all the effort, equipment, and capital invested in the old system would quickly be made obsolete by the new"--seemingly without crediting Edison with an awareness that if his efforts could produce that result, it also could happen at the hands of others. For discussion of other historian's explanations of Edison's behavior, see David (1987), pp. 29-30.

36 J. Hughes (1986), pp. 193-4. This account bears indications of the influence of Josephson's (1959, p. 349) view that the "whole dreadful controversy can be attributed only to an extreme bitterness of feeling towards his [Edison's] opponents that completely overbalanced his judgement."

the conduct of the "Battle of the Systems" has colored modern appraisals of Edison the man, and more with the proper appreciation of how his strategic goals, and technological developments outside his immediate control shaped the decisions that Edison and his close associates made and implemented in this episode.

Initial safety concerns about the hazards of a house-to-house incandescent lighting system based on a.c. notwithstanding, practical experience in the industry soon revealed that grounding and insulation methods worked quite reliably.[37] It is therefore difficult to credit the view that Edison continued to pronounce a.c. "unsafe" because he remained ill-informed, or foolishly obstinate. Presumably, he had a purpose in going on with the campaign until 1889, but not continuing it beyond that year. Indeed, the existence of some deeper, goal-directed strategy has been supposed even by some historians who have expressed disapproval of the tactics used.[38]

An immediate purpose is not too hard to find. The objective of the animal execution demonstrations, the initiation of the electric chair, and the volumes of published material detailing the actual and potential hazards of a.c. was not, however, that of gaining sales in contested markets by influencing the choices made by purchasers of lighting systems, or their ultimate customers. Rather, the proximate goal was the creation of a climate of popular opinion in which further inroads by a.c. systems in the Edison companies' share of the incandescent lighting market might be prevented by legislative restraint, ostensibly justified on grounds that the public's safety needed protection. The gruesome promotional campaign was designed to persuade lawmakers to limit electric circuits to 800 volts, thereby stripping the rival system of the source of its transmission cost advantage. In at least two states, Ohio and Virginia, these efforts came closer to succeeding than in New York, where, having adopted the electric chair, the legislature balked at the intended sequel.[39]

If the events just recounted cannot be dismissed merely as the spiteful reactions of a short-sighted inventor who was being left behind in the onward

---

37 Woodbury (1960), pp. 172-73; Josephson (1959), p. 345.

38 See, for example McDonald (1962), p.62; T. Hughes (1983), pp. 107-09.

39 T. Hughes (1983), p. 108.

march of technology, or simply as an unseemly tussle over marginal market territories between the respective suppliers of d.c. and a.c. lighting equipment, how should they be understood? What was Edison's underlying purpose, and, when was it formed? If Edison was quite sanguine in the face of the commercial initiation of Westinghouse's lighting system in Buffalo, N. Y., as his communication with Johnson in November of 1886 seems to indicate,[40] what had triggered the launching during the following year of the "sordid" propaganda campaign aimed at crippling the a.c. technology with safety legislation? And, having once committed himself to this mode of competition, why did Edison discontinue it after 1889 and quickly withdraw from the electricity supply and equipment business?

Although definitive answers to these questions are elusive, Edison's actions can be construed as an economically rational, albeit cynical response on the part of an inventor-entrepreneur whose long-term plan to be the sole-sponsor of a "universal" electrical supply system suddenly had gone awry. On this view, far from being foolishly obtuse, Edison was more likely than any of the interested parties to have perceived during 1887 that the world in which he was operating had been abruptly altered; specifically, that a number of unexpected and serious blows had been dealt to his previous hopes of profiting greatly from control of the key technological components of a system that could supply energy for public arc-lamps, residential and commercial lighting, industrial power, and traction.

This sudden reversal of his fortunes in the direct current electricity business came from the concatenation of two fundamentally unrelated sets of developments. One was a clustering of induced technological innovations that opened the way for polyphase alternating current to become the basis of an alternative, fully competitive universal electricity supply system. The other, and in our view, the less significant of the two, was the strains experienced in Edison's personal financial position, and that of the companies in which he was most directly involved. The following section focuses upon the key technological developments which precipitated a course of action well-designed to salvage for Edison and his immediate associates as much as was possible before he would exit from the industry to embark upon other ventures.

---

40 See Josephson (1959), p. 346.

3. Polyphase A.C. and the Advent and Significance of the Rotary Converter

The events which marked the episode of intense and open rivalry between the proponents of direct and alternating current known as the "Battle of the Systems" become fully comprehensible only when they are seen to have been precipitated by a fundamental transformation of the underlying techno-economic situation in which a "gateway" innovation played a central part. That disruption of the status quo ante occurred in 1887-1888.

The original situation of "stand-off", which had left d.c. and a.c. in 1885-86 with respective markets within which each held a clear, technically-derived advantage, was in reality a condition of transient equilibrium, a temporary balance soon disturbed by the realization of a differentially faster rate of technological advance along the a.c. system trajectory. Thomas Hughes[41] has suggested that this quickened rate of technical progress should be viewed not only as having been induced, but, somehow as entirely foreseeable. On his interpretation, electrical engineers, having used a.c. to overcome the transmission cost constraints that encumbered the expansion of d.c. distribution networks, quickly encountered numerous obstacles in getting an a.c. electricity supply system to emulate the functions which existing d.c. systems were capable of performing--particularly, the provision of power to secondary motors. Overcoming those obstacles therefore, logically, formed the next engineering task. How foreseeable were the break-throughs that came in 1887-1888 remains another question, however.

While it is appropriate to view such constraints as having served as "focusing devices" (to use Rosenberg's (1969) phrase) which directed engineering efforts towards projects which were more likely to have high commercial payoff if the technical problems could be overcome, the mere existence of the technological drawbacks in the early a.c. system could not have signalled the likelihood that engineers would soon manage to actually solve those problems. In the event, the solutions produced were less than complete, but the train of reactions set in motion by their advent proved decisive in displacing the direct current technology from its prior position of dominance.

In the alternating current field, three new and interrelated innovations abruptly removed the main disadvantages that had formerly restrained a.c.

---

41 See T. Hughes (1983), pp.106-11.

supply companies from penetrating the d.c. lighting and power markets: (1) the induction motor, (2) the a.c. meter, and (3) the rotary converter.

Between 1885 and 1888 Nikola Tesla, in the United States, Galileo Ferraris of Italy, and Michael Osipowitch von Dolivo-Dobrowolsky, in Germany, each had discovered that two alternating currents differing in phase by 90 degrees, or a three alternating current differing in phase by 120 degrees, could be used to rotate a magnetic field; that by placing in such a field a pivoted bar of iron, or a drum of laminated iron on which are wound self-closed coils of copper wire, one could construct a two- or three-phase induction motor.[42] Invention of the motor led directly to the development of a special polyphase generator and system of distribution. Tesla's first two patents on the alternating-current motor were filed in October of 1887, and three more patents pertaining to the system were filed in November. The Westinghouse Electric Co. acquired these patents in July, 1888, by which time Westinghouse already had persuaded Ferraris to permit it to file an American patent on the rotating magnetic field motor which he claimed he had built three years before.[43]

Coincidentally, in April 1888, Oliver B. Shallenberger, an engineer working for Westinghouse came upon the polyphase induction phenomenon and applied it to solve the problem of designing an effective meter for alternating current; by June he had shown that such meters could be built in quantity. Hence, by 1889, Westinghouse Electric Company was able to put into commercial production a meter which enhanced the scope for a.c. in central station lighting operations, and the technological basis for penetrating portions of the electric power market.[44]

---

42 See Fleming (1921), pp. 146-7 for details of motor development and their construction.

43 T. Hughes (1983), pp. 115-17 gives full details on the filing, issuance and acquisition dates of the key Tesla patents, and others.

44 See Prout (1921), pp. 128-29. Development of an a.c. meter was regarded as sufficiently important that Westinghouse himself filed patents on one such device, in June 1887, and on another, in October 1887, developed jointly with one of his engineers, Phillip Lange. Patents were issued for these meters in May 1888, but by then they had been superceded by Shallenberger's design.

While the a.c. induction motor and meter posed an obvious threat, they did not immediately give rise to commercial challenges to Edison's manufacturing companies especially in urban markets where d.c. power supply was already established. Despite the greater than fifty percent market share that a.c. had achieved in the field of lighting by 1891, there was a considerable amount of momentum built into the growth of d.c. supply systems. Two forces were behind this momentum. First, where a local electric supply system did constitute a natural monopoly and there already existed an installed d.c. generating and distribution system, the unamortized investment in d.c. plant was so large that it discouraged replacement with polyphase a.c.; by adopting a mixed system the local utility would lose the scale advantages associated with a single system.[45] Second, as long as there was an important portion of the urban market held firmly by d.c.--namely, the market for traction power, and to a lesser degree industrial electrolysis, as distinquished from that for secondary-motor power for elevators, factories and other such applications--the local d.c. central stations meeting this (traction) demand would necessarily also supply the other electricity needs in the area; rivals could not penetrate the market because the d.c. system with its load balancing advantage could under-price them. Hence unless some "gateway," or interface technology emerged, a partial commitment to perpetuation of the d.c. technology would remain for some time to come.

Viewed from this perspective, it was another technical innovation, part of a new "gateway" technology coincident to and dependent on the development of polyphase a.c., that was critical in delivering the coup de grace to a comprehensive electricity supply system based entirely on a d.c. technology.[46] The rotary convertor was a device which combined an a.c. induction motor with a d.c. dynamo to make possible the connection of high voltage a.c. transmission lines to d.c. distribution networks. A former Edison employee, Charles S.

---

45 T. Hughes (1983), p. 120, notes also that the electrical equipment manufacturing companies "remained partially committed to direct current" by virtue of their specialized facilities, and patent positions, as well as the experience and expertise they had built up in that field.

46 The role of "gateway" innovations in network technology evolution is explicitly considered with reference to the rotary converter by David with Bunn (1987), and is discussed more generally in regard to standardization policy issues by David (1986b).

Bradley, who already had applied for a patent on a polyphase generator and a synchronous motor in March, 1887 filed two months later (May) for a rotary converter patent, which was awarded to him in October 1888.[47] The basic idea for the rotary converter had its origins in the "motor-generator", which simply coupled the drive shaft of an (a.c.) induction motor to a d.c. dynamo, mounting the two assemblies on the same bed-plate. Instead of using two machines to perform the current conversion, the rotary converter offered greater efficiency and reduced cost (and size) by putting both sets of windings on a single armature.[48]

The major significance of the rotary converter lay in the fact that it enabled the old d.c. central station and traction distribution networks to be coupled with new long distance high voltage a.c. transmission mains. These gateway devices thereby made possible the formation of a more flexible "hybrid" system, the advantages of which were recognized in the electrical engineering press as early as 1887.[49] Referring back to the terminological distinctions that have already been introduced (above, in section 1), it may be noted that the rotary converter was to have the functional effect of creating both "compatible complements" and "compatible substitutes" within the emerging universal electricity supply system. The device enabled existing local (three-wire) d.c. distribution networks for lighting and, and also d.c. electric street-car and railway plants, to be used as the complements of longer-distance a.c. transmission lines, and a.c. generators; similarly, the converter permitted the less costly a.c. transmission lines to become complementary to existing d.c. generating stations. Furthermore, by allowing d.c. motors as well as polyphase induction motors (from 1891 onward) to be operated off a.c. electricity supply mains, the

---

[47] Prout (1921), p. 131; Passer (1972), pp.300-01; T. Hughes (1983), p. 118. The courts eventually ruled against the patent application by Bradley for the generator and synchronous motor, on the ground that Tesla's patents were fuller and more complete.

[48] See Encyclopaedia Britannica (1910-11), v.27, pp.137ff. for further details.

[49] See Pfrannkuche (1887) for discussion of an "all purpose" system involving d.c. dynamos and a.c. transmission.

rotary converter created compatible substitutes.[50] The two types of motors were not rendered perfectly compatible substitutes, however; although more-or-less interchangeable in some applications, the performance characteristics of a.c. and d.c. motors differed inherently in ways that, despite the subsequent efforts of electrical engineering firms, left each motor type better suited for some purposes than for others.[51]

These implications of the rotary converter made their practical and commercial impact felt in the United States, not instantaneously, but, progressively over the course of the 1890's as they induced and interacted with other technical and organizational developments in the industry.

By the early 1890s Bradley had set up his own factory to produce the converter in Yonkers, N.Y., a plant that would soon be acquired along with the patent by the General Electric Company. The potential profitability of the business of supplying rotary converters likewise drew the Westinghouse Electric Company

---

[50] The development of technically efficient and commercially successful a.c. motors for industrial applications did not follow immediately from Telsa's patents, or his work for Westinghouse during 1888-89. Indeed, when Telsa left the Westinghouse Co. late in 1889 the company had manufactured only one type of induction motor--a 1/6 HP, single-phase fan motor; the design for a 3-HP two-phase motor, developed in 1890 by Charles F. Scott, a Westinghouse engineer who had worked with Telsa, became the basis for the first polyphase a.c. motor to be commercially installed in the U.S.--when it was applied to drive a coal-cutting machine in the Monongahela, Pa. coal region. See Passer (1972), pp. 256-57, 277-79; Kline (1987), pp. 291-297, discusses the reasons for the failure of Telsa and Scott's motor designs, and the development abroad, during 1890-1892--by engineers working for the Allgemeine Elektrizitats-Gesellschaft (AEG) in Berlin, and the Maschinenfabrik Oerlikon of Zurich (M.O. v. Dolivo-Dobrowolsky and Charles E. L. Brown, respectively)--of the three-phase motor that became the prototype of subsequent induction motors in Europe and the U.S.

[51] Because a.c. (induction) motors had no need to employ commutators, they did not give rise to the problem of sparking at the commutator-brushes which rendered d.c. motors unsatisfactory for applications in environments--such as coal mines, flour mills, and cotton-textile mills--where the danger of dust explosions was serious. On the other hand, induction motors initially suffered from low starting torque, and, even after that defect was rectified by design improvements (See Kline (1987), pp.297-300), the ability of d.c. motors to deliver high torque at low r.p.m. made them much to be preferred for electric traction work and other applicatons where continuously variable speed control was a desideratum.

into further development work on such devices.[52] Converters were also developed to couple existing single-phase a.c. with the newer and more efficient polyphase technology. In fact, by the middle of the 1890s there would exist devices to convert in any direction. Conversion of d.c. to polyphase a.c. from that point onward proved an immediately attractive application in some metropolitan centers, such as Chicago, where Samuel Insull saw it would permit raising the load factors on existing d.c plant by transmitting current over a much more extensive area in which the load was more diverse .[53]

What had happened during 1887-88, in essence, was the dramatic appearance of a new technological variant--the polyphase a.c. system--induced by the opportunities inherent in the limitations of d.c. and single-phase a.c. as bases for a "universal" electricity supply system. It was the polyphase a.c. technology that would diffuse rapidly, becoming the de facto network standard for electricity generation and transmission in the United States, penetrating the core urban markets for electric light and power, and thereby realizing the greater economies of scale which fostered the emergence of extensive "natural monopolies" in the electric supply industry.

Although Edison may not have foreseen the whole evolution of the network technology that would be created to exploit the transmission cost advantages of alternating current, his course of action from 1887 onwards can be construed as reflecting an astute grasp of the precarious, unstable nature of a competitive situation that was unfolding with unexpected speed from the successful

---

52 Lamme (1926), Ch. 6, recounts his early work for Westinghouse on rotary converters.

53 See McDonald (1962), pp. 69-70, on the work of Chicago Edison's chief engineer, Louis Ferguson, circa 1894, who installed rotary converters at both the generating and local distribution ends --using polyphase a.c. for transmission only. On the technology of rotary (or, in some English usage, "rotatory") converters, as well as rotary transformers (used in changing the voltage of direct current), and the Ferranti rectifier (used in transforming an alternating single-phase current into a direct pulsating current for arc lighting), see the article on "Transformers" contributed by J. A. Fleming in Encyclopedia Britannica (1910-1911), vol. 27, pp. 178-79. More generally, on the significance of converters, see T. Hughes, (1983), pp. 121-2.

experiments with polyphase a.c. motors.[54] That perception would have reinforced whatever other considerations might have disposed him to leave the electricity industry, instead of girding himself to participate in an inventive race against the emerging a.c.-based universal system.[55]

As Thomas Hughes has emphasized, Edison's talents inclined him toward the invention of devices that were interrelated within a system context, and he naturally preferred to work on components whose improvement would lead to enhanced performance and value throughout a system whose parameters he could control--and thereby draw the greater profit from.[56] Yet, in the circumstances of the electricity supply industry during 1887-1888, reasonable expectations of extracting significant rents on any incremental, strictly d.c.-compatible inventions had largely been vitiated by the developments leading to Bradley's rotary converter. Furthermore, the d.c. system elaborated under Edison's sponsorship was now at a disadvantage; additional research and development resources would have to be devoted to reducing distribution costs, just for it to be able hold its own in competition with polyphase a.c. systems. What profit to an inventor undertaking that uncertain mission, when the limiting value of the improvement would be that set by the cost of the rotary converters installed to transform d.c. to a.c., and/or a.c. back to d.c. for local distribution and traction uses? Edison

---

54 This is the interpretation developed by David (1987), upon which we draw for the following discussion.

55 Edison, in an 1908 encounter with George Stanley, the son of William Stanley, whose improvement on the Gaulard Gibbs patent had been the basis for Westinghouse's a.c. system, is said to have remarked: "Oh, by the way, tell your father I was wrong." Josephson (1959, p. 349) in reporting this interprets Edison to have thereby acknowledged that he had made "his greatest blunder" by not following Westinghouse into the a.c. technology. Putting aside the issue of whether Edison had made a blunder or a justifiable decision, it remains quite unclear exactly what Edison felt he had been "wrong" about. Possibly it was his mistaken expectation, in 1886, that a great deal of experimenting and a long period of further practical development would be needed before a system based on a.c. could be brought to the point of challenging his own.

56 See T. Hughes (1983), pp.22-23.

had more tempting projects in which to engage his laboratory staff and his own inventive genius.[57]

It is interesting to consider that this inventive penchant on the part of Edison for "system integrity" may have had a bearing on Bradley's decision to leave Edison's employ and to develop his rotary converter as the basis of an independent commercial venture. That Bradley also did not offer this invention to Edison's competitor, Westinghouse, and was not approached by the latter to sell the U.S. patent rights, is even more remarkable. There were at the time engineers in the Westinghouse organization who recognized the significance of the rotary converter as a remedy for the inability of the company's (single phase) a.c. system to serve traction-power users of electricity. But Westinghouse himself does not seem to have lent much encouragement to research and development efforts in that direction. Benjamin G. Lamme, later to rise to Chief Engineer at Westinghouse Electric Co., was working on the conversion problem in the late 1880's at the company's East Pittsburgh plant. But as he subsequently recalled:[58]

> "Strangely enough, it appeared to me that Mr. Westinghouse never took any strong interest in the rotary converter as affording a means for extending the field of direct-current traction, and yet this has been possibly the greatest single step in overcoming the early limitations of the 600 volt system. He did not ask me many questions regarding rotary-converter development and operation, as he did with other developments."

Westinghouse's biographer, Prout (1921) found this hard to explain. It raised the possibility of the presence of "a small and passing dark spot in [Westinghouse's] imagination just there." But an alternative rationale might be found in the suggestion that George Westinghouse was no less keen than Thomas Edison to be the sponsor of a universal electricity supply system, based on a single form of current. From the viewpoint of either of the rival sponsors, then, a gateway technology like the rotary converter would not have

---

57 Thus, it is not surprising that after the spring of 1887 Edison should have allowed himself to become increasingly pre-occupied with renewed experimental work on his phonograph--entering an inventive race against workers in the laboratory of Alexander Bell who had undertaken to improve upon the original Edison device of 1877. See Josephson (1959),pp.317-31.

58 The following passage is quoted by Prout (1921), p. 131.

appeared as such "good news" so long as they hoped to be able to capture most of the rents rather than having to share the profits of a "hybrid" system which made use of patented components from both camps.[59]

The unorthodox and rather desperate tactics adopted by the "West Orange Gang" in their "safety campaign" against Westinghouse takes on a different appearance when set in this context, especially their focus upon invoking some regulatory intervention to deprive alternating current-based systems of the transmission advantages deriving from use of high voltages. The campaign looks more like a temporary "holding operation," meant to delay the competition so as to permit Edison to stage a more orderly and profitable exit from the industry, and less like a serious counter-offensive. Had it been Edison's hope and intention to permanently cripple the competitive system of electricity supply, it would hardly have made sense for him to accelerate his withdrawal from the business of manufacturing and selling the key components of the d.c. system that would be left in command of the field. Yet, by 1888, that is precisely what he was proceeding to do.

Edison's decision to get out of the electricity business was significant, because it would lead shortly to the disappearance from the United States' electrical manufacturing industry of a commercial sponsor having proprietary interests exclusively in the d.c. technology.[60] It had its roots not only in the burst of polyphase a.c. developments just reviewed, but in the evolution of Edison's relationship with the holding company (Edison Electric Light) in charge of his lighting patents and the various entities such as Edison Lamp, and (Edison) Machine Works, that actually manufactured the components of the system and serviced them. Recall that by the mid-1880s Edison and his immediate associates had little stake in, and less control of the holding

---

59 Eventually, Westinghouse did express pleasure with the way the rotary converter developed during the period 1894-99, as Lamme reported. See Prout (1921), p.133. By that time, however, he no longer was worried about competition from an integral supply system based upon direct current; Edison had left the industry and, from 1896 onward, General Electric and Westinghouse had entered a patent pooling agreement--creating the Patent Control Board--to share the rents on both the a.c. and d.c. technologies.

60 See Josephson (1959), pp. 350-66, and McDonald (1962), pp. 30-39 for a fuller account of the formation of the Edison General Electric Co., and then, the General Electric Co.

company that drew royalties on the use of the lighting patents by central station companies and other companies set up to license the construction of isolated lighting plants; whereas, Edison remained the principal owner of the factories, from which he and his immediate associates had been drawing their main income. During 1886-1888, however, the precarious financial situation of the latter group of enterprises came to be perceived as tremendously burdensome to their owners. Much of the equipment supplied previously by the manufacturers to central stations had been paid for with hard-to-negotiate shares in those fledgling enterprises, resulting in severe cash flow problems for the Edison concerns. With the recovery from the business recession of 1885, electrical equipment orders were coming in faster than the factories had resources to produce and deliver. But, it was proving difficult to finance the expansion on short-term credit, and to solve this problem by permitting the Morgan banking group (who already dominated Edison Electric Light) to extend their control over the manufacturing enterprises in exchange for long-term loans was hardly an appealing prospect.[61]

Consequently, the return of Henry Villard from Germany in 1886, bearing a commission from the Deutsche Bank to negotiate with Drexel, Morgan and Company about the acquisition of holdings in American businesses, came at a most fortuitous moment. Edison was much relieved to be presented soon thereafter with Villard's proposed plan for the consolidation of all the Edison-related enterprises (including the patent-holding company, and Sprague Electric Railway and Motor Company) into one new corporation with backing from the Deutsche Bank, the Allgemeine Elektrizitats Gesellschaft, and the firm of Siemens and Halske, of Berlin. The inventor saw in this a welcome opportunity to both extricate himself from the worries and distractions of managerial and

---

[61] Thus, at the same time that Edison was fretting over the possibility that his over-expanded manufacturing businesses might be forced into bankruptcy, he was also sending his financially skillful young associate, Samuel Insull, off to borrow the extra sums needed for the West Orange, N.J. laboratory's expanding program of non-electrical researches. See Josephson (1959), p. 340; McDonald (1962), p. 38.

financial responsibility for the manufacturing business, and raise sufficient capital to place his laboratory on firmer financial foundations.[62]

In this way the Edison General Electric Co. came to be organized in January, 1889. Within a few months the consolidations were formally effected, and Edison himself no longer had direct influence in the running of the manufacturing side of the business. By 1890 he had largely completed the liquidation of his remaining 10 per cent shareholding in the new company, was taking no active role on its board of directors, and was writing to ask Villard not to oppose his "retirement from the lighting business, which will enable me to enter into fresh and congenial fields of work."[63] The propaganda war against the Westinghouse a.c. system, which had been brought to its peak in the midst of the consolidation negotiations, was rapidly wound down in 1889. It would seem to have served the real purpose of supporting the perceived value of the Edison enterprises which were at the time wholly committed to manufacturing the components of a d.c.-based electric light and power system, and thereby improving the terms on which Edison and his close associates were able to "cash out."

The electrical manufacturing business in the United States thus came to be dominated by two large firms from 1892 onward, but, the industry also had become essentially homogeneous with regard to the basic formulation of its technology. By 1893, the General Electric and the Westinghouse Electric and Manufacturing Companies both were marketing some version of a polyphase alternating current system, and both had entered the profitable business of manufacturing rotary converters.[64] The era of rivalry between commercial sponsors of technologically distinct systems in the United States' electrical supply industry was brought to a close within six years of its commencement.

Yet, the question of the superiority of one form of current over the other remained unresolved within the engineering community. Whether direct or

---

62 Josephson (1959), pp. 351-3. On Villard's role, see also, McDonald (1962), pp.39-40, and T.Hughes (1983), pp. 76-77.

63 Josephson (1959), p.361, quoting a letter dated February 8, 1890.

64 Passer (1972), pp. 298-303; Sharlin (1963), pp. 187-88. Bradley's patents and facilities were bought up by General Electric and the Westinghouse Electric Company developed their own rotary convertors over the same period.

alternating current was to be generated by the hydroelectric power project being undertaken at the Niagara Falls was still very much an open question during 1892.[65] The proponents of d.c. at Niagara had argued that for conditions of varying load, as was the case in a lighting system , d.c. was much the more efficient of the two.[66] While this may have held true in 1890, following the 1891 demonstration by Oscar Muller and the Swiss firm of Boveri and Co. that polyphase current could be transmitted the 110 miles from Lauffen on the upper Neckar River to Frankfurt-am-Main, Germany,[67] and the equally impressive Westinghouse Electric Company polyphase system exposition (including the rotary converter) at the Chicago World's Fair in 1893, it was evident that lighting was no longer the only factor to consider when discussing the load and efficiency characteristics of the a.c. and d.c. system variants.[68] With the extension of a.c. to power users, and to traction users as a result of the invention of the rotary converter, the decision between a.c. and d.c. came down to one of which could distribute power over a distance most efficiently and cheaply. And, as the distance from Niagara to the nearest concentration of customers, in Buffalo, N.Y., was 20 miles, a.c. could reduce the loss of power on transmission lines to a far greater extent than was possible with d.c. In 1893, both the Westinghouse Electric Company and the newly formed General Electric Company submitted plans to the Cataract Construction Company (which was pioneering the Niagara development), specifying an a.c. system consisting of generators, transformers, and transmission lines.[69]

---

65 Fleming (1921),pp. 238-9.

66 "One engineer estimated that under such conditions, the d.c. system would be twice as efficient as an a.c. system." Sharlin (1963), p.200 , quoting from a statement in 1900 by the British electrical engineering authority, J. A. Fleming.

67 See Landes (1969), p. 286, and Fleming (1921), pp. 238 ff. for technical details.

68 On Westinghouse exhibit at The Columbian Exposition in 1893, see, e.g., Prout (1921), pp. 134-40; Sharlin (1963), pp. 206-211.

69 See Sharlin (1963), pp. 195-210 for discussion of the choice of technology at Niagara.

The development of the rotary converter seems to have played a quite direct part in George Westinghouse's own commitment to implementing the concept of a universal electrical supply system based upon a.c. as the solution to the problem posed by the Niagara project. Westinghouse intially had thought of transmitting power from the hydrogeneration site at Niagara to Buffalo by pneumatic (!) means--which is perhaps more understandable when one recalls that he had been working on compressed air systems (for train-brakes) for two decades. But, according to Prout (1921):[70]

> "the success of the Telluride plant [built in 1890-91 in Telluride, Colorado, to carry power 3 miles at 3000v., single-phase a.c.] and the progress at Pittsburgh in the autumn of 1892, particularly in the development of the rotary converter and of two-phase motors, definitely changed his mind."

Beginning in 1896, central stations began being converted into sub-stations hooked up to a.c. transmission lines, and by 1898 a constant current transformer had been developed to make possible the linking up of arc lighting distribution networks with a.c. transmission lines.[71] Hence, the flexibility and capabilities of this new coupling technology led to a rapid diffusion of the a.c. polyphase technology and the integration of smaller urban electricity supply systems into larger networks which eventually formed regional grids.[72]

Once the systems competition had tipped in this direction, lighting plant also came to be replaced with a.c. technology as the previous d.c. distribution networks wore out; other things being equal, transformers stepping-down high voltage a.c. were a much simpler and cheaper technology for lighting purposes than the use of rotary convertors to feed local d.c. distribution networks from high voltage a.c. transmission lines.[73]

Convertors continued to be employed well into the 20th century, however, most

---

70 Prout (1921), pp.130-31; 145 for following quotation.

71 National Electrical Manufacturers Association (1946), p. 74.

72 See Byatt (1979), p. 114; Bowers (1982), p. 162; Lardner (1903) for discussion of regional grid development.

73 See Byatt (1979), p. 76.

notably in the traction field where d.c. remained the current preferred at the point of consumption.[74]

A de facto standard, in the form of alternating current as the basis of a "universal" electricity supply system had emerged by the 1920s, both in the United States and abroad. While diffusion data for the 1890s is not available, the figures that can be assembled show that in America the fraction of central station generating capacity accounted for by a.c. rose from 69 percent in 1902 to 95 percent in 1917. Moreover, one can discern in this the large role played by the rotary converter. This appears from the absolute rise in rotary converter capacity installed and also from the fact that the d.c. share of end-use capacity fell far less sharply than its share in generating capacity. Indeed, the former remained essentially unchanged after 1907.[75] At the engineering level, therefore, the "battle of the systems" did not end with the capitulation and withdrawal of one of the contenders, as the Edison-Westinghouse business rivalry had done. The technological denouement has been described by Thomas Parke Hughes[76] as a peaceful resolution to the conflict: no outright defeat for d.c., rather, a graceful and apparently efficient absorption within a transitional "mixed" system, prior to its gradual disappearance from the American electrical scene.

The perspective of hindsight may impart to this story an impression of inevitability, and even a supposition of optimality, both of which should be resisted. Meaningful global evaluations of efficiency are difficult if not impossible to make between alternative technological systems whose influences ramify so widely and are so profound that they are capable of utterly transforming the economic and social environments into which they have been

---

74 See Byatt (1979), ch. 4; Sharlin (1963), pp. 185,188.

75 See David (1987), Table 1.

76 T.Hughes' (1983) pp. 120-21 succinct formulation bears quotation: "Because 'the battle of the systems' had become far more complicated than a technical problem awaiting a simple technical solution, it ended without the dramatic vanquishing of one system by the other, or a revolutionary transition from one paradigm to another. The conflict was resolved by synthesis, by a combination of coupling and merging. The coupling took place on the technical level; the merging, on the institutional level."

introduced.[77] As a guard against the strong temptation to suppose in matters of technological advance that "what is, ought to have been", it is always useful to at least notice the existence of other paths that were not taken.

From the technical journals and magazines of the period it is apparent that no immediate consensus emerged on the engineering merits of the two currents; there were well-respected inventors and scientists, many of whom were founders of the industry, who would not testify to the technical superiority of alternating current.[78] For, the d.c. technology also showed itself quite capable of being further elaborated, in directions that both heightened its special advantages and broadened the range of conditions under which it was economically competitive.

As was the case with a.c., the possibility of lowering the cost of the d.c. systems by raising the voltage was being actively explored. As early as 1883 Charles F. Brush had attempted a high voltage d.c. system that could more fully utilize fixed generating capacity and increase the radius of profitable transmission, by using "accumulators" (storage batteries) to handle some of the peak lighting load and to accomplish the reduction of voltage for local distribution. Due to a combination of problems associated with the battery technology available at the time, the dangers of operating a high potential system having dynamos wired in series, and the usual run of financial difficulties, this particular project never took off and the concept was not pursued further in the United States.[79]

---

[77] This is a difficulty with which efforts to quantify the effects of major transport innovations, such as the canal and railroad systems, have had to contend (see, e.g., David (1975), Ch. 6); We are grateful to Edward Constant, who has reminded us of its relevance in the present connection.

[78] Throughout the "Battle of the Systems" electrical engineers debated the advantages and disadvantages of the variant systems in the pages of the Electrical World (see, e.g. Duncan (1888) and February 26, 1887; January 21, 1888; March 31, 1888). In addition to Lord Kelvin, and Dr. Werner von Siemens, two of the more distinguished industry personalities who argued on behalf of direct current until late in the 19th century, there were many other engineers continuing to debate the merits of the two systems at the turn of the century; see, e.g., Barstow (1901) and Scott (1901).

[79] Stanley (1912), p. 565.

Abroad, however, the English from the mid-1880s onwards attempted to implement a variety of h.v.d.c. electricity supply schemes.[80] Several innovations introduced late in the 1880s did prove successful. Notable among these was the "Oxford system," an approach first employed in 1889, which transmitted high voltage d.c. to substations where it was "bumped down" to usable voltage levels via either a battery arrangement or a direct current transformer.[81] On balance, the major advantages of the h.v.d.c.-battery technology lay in ensuring continuity of supply when generating plants failed or were shut down for maintenance, and in reducing the amount of fixed capacity required to meet peak loads on the system.[82] In Britain, then, the story unfolded along lines very different from those in the United States; there, the "competition of the currents" was tipped during the 1890s towards d.c. by the possibility of using accumulators in combination with high voltage.[83] Concurrently, on the European Continent,

---

80 The voltages these involved were much lower than those in modern high voltage direct current systems, which recently have re-emerged as a technological area of active research and development interest to the electricity industry in the United States. See, e.g., Alvarado and Lansetter (1984), Weeks (1981), Zorpette (1985).

81 See Parsons (1940), ch. IV, pp. 52-70, for discussion of early h.v.d.c. systems in England. According to the Encyclopedia Britannica (1910-1911), vol.9, p. 196, the "Oxford system" was distinguished by the use of "continuous" (direct) current transformers to accomplish the drop in voltage from the 1000-2000 v. range at which the current was transmitted from the generating stations to the 100-150 v. supplied by distributing mains to users. Although 3000 v. came to be regarded as the practical limiting voltage for individual d.c. dynamos (due to the problems of sparking at the commutator brushes when running at faster speeds) two or more such machines could be wired in series in order to secure much greater voltages for purposes of transmission. In France, for example, on the Thury direct current system energy was transmitted a distance of 115 miles (between Moutiers and Lyons) at voltages upwards of 45,000 v., using 4 groups of dynamos in series, each group consisting of 4 machines in series. See Encyclopedia Britannica (1910-1911), vol. 8, p.778.

82 Byatt (1979), p. 100. One additional advantage of the "Oxford system" was that existing d.c. arc lamps for street lighting could be worked off the high voltage mains in sets of 20 to 40. See Encyclopedia Britannica (1910-1911), vol.9, p. 196. See also Lineff (1888).

83 Landes (1969), p. 285-86 summarizes this with the statement that "The two systems competed fiercely in Britain for many years. In the long run, however, victory lay with centralized generators and long-distance

M.Thury was developing a system of transmitting direct current at very high voltages from constant current generators worked in series, and commonly coupled mechanically in pairs or larger groups driven by a single prime mover. This offered advantages in easier line insulation than was required at half the voltage with alternating current, and removal of difficulties of line inductance and capacity encountered in high voltage a.c. transmission. High voltage constant-current plant lent itself to greater ease of operation in emergencies (over a grounded circuit, for example) and permitted the design of comparatively simple and inexpensive switchboard arrangements. Notwithstanding the fact that the direct current generators used in this system were relatively expensive and their individual output was inconveniently small for large transmission work, around 1910 even a contemporary American authority gave it as his view that[84]

> "The possibilities of improvement in the system have by no means been worked out, and although it has been overshadowed by the enormous growth of polyphase transmission it must still be considered seriously."

Today it is known that very high voltage d.c. transmission lines are free from the problems of reactance, and "corona effects" in a.c. long-lines with which scientists and electrical engineers from Tesla onward had to deal.[85] It is high voltage per se which makes for cost-savings in transmission work, and, in the 1890s, when the transformer provided a comparatively cheap way to raise

---

transmission." Although the latter is indisputable, there is a point in noticing that the "victory" was not an indigenous evolutionary outcome. It came in the 1920s, with the transplantation to Britain of the electric utility system technology which had become the dominant engineering style in the United States, and so did not represent an independent experiment; indeed, had the a.c.-d.c. rivalry not been resolved in the American market so much earlier, allowing time for much improvement in the design and actual operation of electricity networks, borrowing technology from the United States might hardly have been so attractive to the British in the 1920s. See Bunn (1987) for further discussion of the British experience.

[84] Louis Bell, "Power Transmission", in Encyclopedia Britannica (1910-1911), vol.22, p.234. Bell was, Chief Engineer of the Electric Power Transmission Department of General Electric Co., and former editor of Electrical World. See also Weeks (1981), pp. 267-71.

[85] On modern "HVDC" technology, see Zorpette (1985).

voltages to very high levels (i.e., by substituting voltage for amperage) the advantage in transmission work was held by alternating current. Now that hydrogeneration and other technology is available for attaining very high voltages with d.c., that current holds the advantage for long, direct-line transmission purposes.[86] The feature of a.c. which proved to be critically advantageous in the development of regional electric power "grids" appears to be the ease with which circuit-breaking is accomplished at the zero-point in the current cycle.[87] Isolating sub-systems that failed, by means of circuit-breakers was recognized to be extremely important once extended networks began to be constructed, but not when the "contest of the currents" was being decided in the U.S. market.

Unlike the battery-using h.v.d.c., the a.c. version of the "universal" electricity system concept inexorably pushed Samuel Insull, and others who pioneered it in the U.S. around the turn of the century, towards "load balancing" as the way to mitigate the wastage imposed by having to build enough generating capacity to meet peak loads. The search for a diversified load over a wider region, with high fixed costs in place, created problems of natural monopolies which would not have existed to the same degree under the h.v.d.c.battery technology. Of course, it must be acknowledged that without strong increasing returns effects via load balancing, a d.c.-battery-d.c.

---

[86] The fact that end-use distributions systems, and high voltage generating plants are based on a.c. requires that conversion be made to h.v.d.c. for "shipment", entailing expensive "terminals" for that purpose. The fixed costs of conversion terminals are quite substantial and increase with their power, giving rise to a schedule of "break-even" distances for the use of h.v.d.c. for transmission purposes, in which the critical distance at which the transmission savings begin to exceed the fixed costs, is a rising function of the wattage. For a 2-terminal installation, the minimum distance at which it becomes cost-effective to ship power by h.v.d.c. lines is now approximately 200 miles. See Zorpette (1985), p. 31.

[87] On the other hand, linkage of a.c. networks that are asynchronous into larger systems continues to be technically difficult. See Zorpette (1985), p.32. One implication of this would seem to be that it was very difficult to build up a large regional a.c. system by ex post integration of smaller local systems, using the advantages of a.c. for transmission lines. A.c. networks therefore had a greater advantage where they could expand "organically" into new territory, as was the situation in the U.S. by comparison with that in Great Britain. See Byatt (1979).

technology might--simply have allowed more leeway to the forces making for too great a degree of diversity in voltages, current and a.c. frequencies.

Such certainly could be said to have been the experience of the industry in Britain, where, the availability of the rotary converter technology notwithstanding, the period 1891-1911 saw nothing comparable to the American pattern of absorption of local d.c. distribution systems for light and power within expanding electric utility networks based on the transmission advantages of alternating current. Rather than revealing the dsyfunctionality of an alternative (h.v.d.c.-battery) electricity supply technology, however, what the British experience serves to show is that the economic impact of a gateway technology--like that of any other technology--may be tightly circumscribed by political, regulatory constraints. The Electric Lighting Acts of 1882 and 1888 not only delayed the building of central stations in Britain, but insured that when electric utilities would develop, they would do so within the confines of many distinct, and geographically small markets which would remain isolated from one another.[88] This legislation followed in the path of previous Acts establishing municipal control of local "natural monopolies", such as tramway enterprises, waterworks, and illuminating gas works--with which the early electric light companies were threatening to compete directly.[89] It stipulated that local authorities would either operate electric utilities or determine whether or not private companies should be permitted to do so; moreover, it prohibited amalgamation of, and cooperation between electric utilities that were organized on this basis. Thus, it may be argued that it was the parochial regulatory context within which the British electricity supply industry was obliged to

---

[88] See Hughes (1962), pp. 30-1; Hughes (1983), p.87-88; Bowers (1982), pp. 155ff. Bunn (1987) examines the historical consequences of this legislation in greater detail.

[89] Since the supply-area radii of the early, d.c. central stations was very restricted by the high costs of copper wire, placing the electricity business under the jurisdication of local municipal authorities, even where boroughs were geographically quite small, as was the case in London, must have appeared to be technologically justified. This is another instance of the way in which an "historical accident"--in the detailed timing of technologically myopic Parliamentary legislative intervention which assigned regulatory authority to the local boroughs, rather than to counties--can produce quite lasting and profound effects upon the course of a network industry's evolution.

operate, which was fundamentally responsible for bringing into existence the multiplicity of isolated markets, each of which electrical equipment suppliers might hope to capture and hold by furnishing with some "customized", idyosyncratic technical setup. The political segmentation of the market into small geographical monopoly preserves both fostered the lack of standardization, and made the d.c.-battery technology economically viable for many municipal utilities.[90]

This is not to say that the course of technological development permitted by the less constrictive regulatory environment in which the industry was able to operate in the United States was one that has proved to be "optimal". Although here the territorial extension of electric utilities' a.c. transmission lines soon outran the reach (even the venal reach of franchise-selling) municipal authorities, the pressures to control the resulting natural monopolies issued in the erection of a regulatory apparatus a the state level. Conflicts between these regulatory jurisdictions have greatly complicated, where they have not blocked the formation and management of the larger regional networks that have subsequently become technologically advantageous even within the parameters of a supply system based upon alternating current--leaving aside one that would make use of modern H.V.D.C. transmission technology.

Such skepticism as we have just entertained about the long-run economic optimality and consequent inevitability of the de facto standard that emerged in the United States is, of course, reinforced by the foregoing recounting of its historical roots in "battle of the systems." Given the urgency of the American utility companies' drive to achieve scale economies in electric supply, their move to the polyphase a.c.-based "universal" system was certainly little affected by weighing the potentialities for long-run technical improvement offered by alternative systems. Just as short-run liquidity considerations had figured prominently in Edison's decisions during the formation of Edison General Electric Co., so, short-run maximization of rents on the existing stock of proprietary technology seems to be the best algorithm descriptive of the course of action pursued by the dominant successor-firms engaged in manufacturing and marketing

---

90 This and other aspects of the British experience are re-examined by Bunn (1987).

electric supply equipment in the U.S. during the years before the First World War.

4. A.C. System Parameter Standardization--"Electrical Engineers do it with Frequency"

The rapid diffusion of polyphase a.c. technology as the basis for universal electricity supply networks in the United States was encouraged by the early move toward the standardization of that system's parameters, and, in particular, by the conscious efforts that American electrical engineers seem to have made to standardize current-generation frequencies. The reason for the special attention accorded to the standardization of frequency--that is, the number of cycles made per unit of time by an alternating current--in comparison with other key system parameters, such as voltage, current, or phase, is to be found in the unavailability of an efficient, low cost gateway technology that could overcome incompatibilities in this dimension. Although frequency converters were developed, they were neither as efficient nor as reliable for this purpose as were rotary converters in permiting d.c. motors to function within an a.c.-based network environment.[91]

The determination of what frequency to specify in designing a.c. plant and equipment was historically dependent on the end-use, or end-uses to which the current was to be put. And, in this way, because it affected the pattern of use made of electric power supplied in the form of alternating current, the rotary convert itself came to play an indirect part of some consequence in shaping the parameters of the evolving a.c. supply system. Its role, however, was not that of promoting complete standardization on a single frequency. As a gateway innovation it had the initial effect of increasing the heterogeneity of end uses; the rotary converter had given a large bloc of power-users with

---

91 Voltage conversions were, of course, available through the use of "transformers", from Gaulard and Gibbs (1882) onward, and formed the basis of the a.c. technology's transmission advantages. As for phase conversion, Charles F. Scott, a Westinghouse engineer, in 1893 invented a method of converting from 2-phase to 3-phase a.c. by special winding and grouping of transformers. See Prout (1921), p.147. The technical problems of connecting polyphase a.c. networks that are standardized with regard to frequency, but are not synchronous, remains so formidable that high voltage d.c. (back-to-back) links are presently employed in moving power between such networks even when they are spatially contiguous. See Zorpette (1985).

frequency requirements that were distinct from those of electric lighting customers access to, and significant influence over the configuration of a.c. supply networks. The consequence, as will be seen, was the creation of a demand for dual standards in regard to frequency, even though, at the same time, the rotary converter was having the effect of tipping the balance towards the establishment of alternating current as the universal system standard.

During the early years of alternating current technology, prior to the advent of the polyphase technology, use of a.c. was restricted to arc and incandescent lighting. Manufacturers of a.c. lighting equipment discovered, somewhat by accident it would seem, that arc and incandescent lighting systems operated well off current with frequency in a range above 100 cycles per second (cps); in the U.S. during the late 1880's the Westinghouse Electric Co.'s lighting utilities initially operated with single-phase a.c. at 133.33 cps, while the Thomson-Houston Company settled on 125 cps. In the course of the former firm's research on the polyphase technology in this period, however, it was found that the new Tesla induction motor, on which so much hope for the development of an a.c. universal system was staked, could not be operated on current drawn from the single-phase generating stations that had already been constructed, or at the high frequencies widely used for electric-lighting work. When it was thus recognized that commercial application of the polyphase technology as a universal system would eventually require the installation of new generation and distribution capacity, the issue arose as to the desirability of choosing a standard frequency. Hence, during the first half of 1891, a point in time at which polyphase power research at the Westinghouse Co. was temporarily suspended due to financial problems, Westinghouse engineers were set to studying what the best frequency standard would be.[92]

George Westinghouse's biographer, Ralph Prout tells us that his subject was[93]

> "probably the first man of strong influence in electric development to realize the importance of adopting and adhering to a standard frequency of alternators. Very early he pointed out to the staff at Pittsburgh that a standard frequency was

---

92 See Prout (1921), pp. 148-49; Passer (1972), p. 279.

93 Prout (1921), p. 148. See also Passer (1972), p. 315.

important in the same sense that a standard railroad gauge is important."

Railroad gauge standardization, it might be noted, was itself a matter that was still "in the air" in this era--the last major step towards guage uniformity east of the Mississippi, the conversion of the southern seaboard railroad system to the standard guage having been accomplished on January 1, 1887, to the accompaniment of much publicity.[94] Whether this precedent was sufficient to commit Westinghouse's organization to a policy of attempting successfully to sponsor a standard frequency for the industry, or whether other, economic calculations entered into the decision, remains difficult if not impossible to determine at this distance from the events.

Looking at the question from the vantage point of various end-uses to which alternating current would be put in homes and businesses, it appears clear that standardization of frequency would have two opposing effects. First, since different end-uses have different frequency requirements, standardization would entail some loss in performance for end-uses where the standard chosen was sub-optimal. On the plus side, however, users might benefit from lower cost equipment if competitive suppliers could, as a result of having standardized designs, use mass production methods and move further down their respective "learning curves" in the manufacturing process. Other types of users, specifically central station utilities and traction companies, would likewise suffer a performance loss if the standard chosen was not that which was optimal for their particular load. They too, however, might benefit from lower cost equipment, and more importantly, reap the tremendous advantage of being able to share power among themselves in an effort to utilize generation capacity most efficiently.

The manufacturers of generating equipment would face a different set of gains and losses from standardization, for these firms would lose the flexibility of modifying standard equipment to meet idyosyncratic requirements of some customers, and would surrender the option to compete for equipment orders and contracts on the basis of performance characteristics which depended directly upon frequency. In return for those potential sacrifices, however, standardization would probably reduce manufacturing costs by permitting

---

[94] See Puffert (1986).

achievement of immediate scale economies, as well as through "learning" effects. In addition, a full-line electrical manufacturer (like Westinghouse) in committing to a standard frequency on end-use lighting equipment, induction motors, rotary converters and generators, would be meeting the technical compatibility requirements for system operators (and possibly also system users) to benefit from "load-balancing" and the cost reductions that could be achieved by raising the load factor. Such considerations, however, are likely to have played more of a role in the convergence by the major U.S. manufacturers of electrical equipment upon uniform frequency standards for the American market, which occurred during the course of the 1890s, than it did in prompting George Westinghouse to set his engineers the task of selecting an optimum frequency. Nonetheless, analogous considerations did underlie the movement toward railroad guage standardization which he cited as a paradigm for that engineering exercise.

The choice of an optimal frequency was not an easy one, for, there were many trade-offs to be considered. Choosing a frequency from the lower end of the spectrum offered advantages over higher frequencies in that (1) smaller, and hence less costly cables could be used for transmission, because at lower frequencies the flow of current is less impeded on the transmission lines; (2) less trouble would be encountered from static charge accompanying the heating of the insulation, and from resonance effects; (3) for a given number of revolutions per minute of a generator, the lower the number of field poles which would be required for that generator--a significant consideration, in that the more poles required on the generator the greater its cost of construction and the more difficult its operation;[95] (4) it was discovered that rotary converters operated best when the alternating current was as close to d.c. as possible, meaning at the lowest possible frequency range.

Detracting from the attractiveness of lower frequencies, however, were at least three disadvantages: (1) at lower frequencies, larger, and hence more costly transformers had to be used to bump the current down to lower voltages; (2) again, given the r.p.m. of the generator, the lower the frequency the lower the speed attainable by induction motors operating off the current[96]; (3) it was

---

95 Passer (1972), p. 280.

96 Passer (1972), p. 280.

found that at a frequency level below approximately 60 cycles per second, arc lights no longer performed adequately.[97]

On balance, it appears that aside from the issue of the operation of arc lights, the engineers would have opted for a frequency towards the lower end of the spectrum, but the need to service existing arc-lighting systems constrained them to a lower bound of 60 cps.[98] This posed a problem in terms of chosing a "single" standard however because the rotary converters then designed required a frequency much below 60 cyles. As a result, latter in 1891, at the close of their investigation, the Westinghouse Co. engineers recommended a dual frequency standard: 60 cycles for plant in markets which were primarily lighting, and 30 cycles in markets where rotary converters could be used "advantageously". This recommendation was approved for implementation in Westinghouse generator designs for the Niagara Project.[99] The choice of 30 cycles had been made over that of 25 cycles, the low-frequency standard that was eventually to emerge, because 30 cycles was both better for incandescent lights and better-suited to induction motors, in that it allowed the latter to operate at speeds 20 percent higher.[100]

General Electric immediately followed their competition's lead in moving to a frequency lower than those which obtained under the single-phase a.c. technology. But, interestingly enough, they did not choose to coordinate with the Westinghouse Company in their choice of either a phase or frequency standard; G.E. chose to go with three-phase a.c. immediately, whereas the Westinghouse Company first went with the two phase variant; and G.E. chose as their two frequencies 66.66 and 36 cycles per second.[101] Continued diversity among frequencies did not end here however. A short while later, prior to submitting their bid for the Niagara project contract in September 1892, G.E. decided that the advantages of having a single as opposed to a two frequency standard were worth pursuing. The company brought out a line of 40 cycle

---

97 Passer(1972), p. 315.

98 Lamme (1918), p. 72.

99 Lamme (1918) pp. 67-70, and Passer (1972), pp. 280-1.

100 Lamme (1918), p. 71.

101 Passer (1972), p. 315.

equipment, hoping that eventually the difficulties this posed for lighting systems could be worked out and this would become the single standard in the industry.[102] In the Far Western United States, where choices were less constrained by importance of accomodating the characteristics of the early rotary converters to tie into existing d.c. distribution systems, and to service large urban markets for traction power, serveral mid-range frequency standards were tried at various locations: 40 cycles in Portland, Oregon, 60 cycles in San Francisco, and 50 cycles in Los Angeles.[103]

The 25 cycle frequency, which along with the 60 cycle frequency, came into prominence throughout the industry in the U.S., was nobody's "optimum". It arose from a negotiated compromise between the Westinghouse Company and the designers hired by the Cataract Construction Company--the organization which was responsible for the large hydroelectric project under construction at Niagara Falls during the mid 1890s. Professor Forbes, the engineer advising the Cataract Company favored a 16.66 cycle installation, because he thought it offered possibilities for commutator-type motors similar to those used with direct current. The Westinghouse Company countered by pointing out that such a frequency was too low for all other purposes and suggested a 33.3 cycle design. This was as close as they could get to their proposed standard of 30 cycles, having been constrained to a speed of 250 revolutions per minute by the promoters of the Niagara scheme. They could not obtain 30 cycles, due to the engineering relationship that exists between the revolutions per minute of a generator, the number of field poles on that generator, and the frequency of the current that is generated: (rpm) x (the number of poles) = (the number of alternations of current per minute); this latter, divided by 120 gives the number of cycles per second. Hence, the Westinghouse Company could use either a 16 pole design to obtain a 33.33 cycle current, or a 12 pole design to obtain a 25 cycle current. Although its engineers pointed out that the rotary converters then being manufactured would operate much more efficiently at the higher of those frequencies, Professor Forbes was apparently obdurate in favoring something

---

102 For a more detailed discussion of the strengths and weaknesses of various frequencies, or possible two frequency standards, see Lamme (1918), pp. 82-83.

103 Ryan (1934), p. 670.

lower. The issue was only resolved by the two parties agreeing to compromise at 25 cycles per second.[104]

Thus, a haphazard combination of technical constraints imposed by performance considerations, taken in conjunction with the limitation of antecedent technologies, the need to adapt to the technical requirements of the new hybrid structure of power demand served by polyphase technology, and the necessity of compromise in order to secure a contract, governed selection of the dual frequency standard followed by the various manufacturers of equipment in the United States from the mid-1890s onward. It appears that, aside from the fact that it was well suited to the work with the earliest rotary converters (which held true also for other frequencies in the lower range), 25 cycles took hold as the low-frequency standard due to the historical accident that it received a great deal of publicity in connection with the electrical engineering side of the Niagara project. The latter, as has been seen, itself lent impetus to the rapid diffusion of polyphase a.c. technology into commercial applications in the U.S. Since induction motors operated best at lower frequencies, and 60 cycles, was the operational lower bound for arc-lighting systems, 60 cycles was the logical choice for the many utilities which serviced a combined public lighting, residential lighting, and power load. While the 40 cycle system proposed by G.E. was described as having excellent features to commend it, at least on engineering grounds, it appears to have emerged on the American scene only after the 25 cycle standard had become entrenched, and thus never attracted much of a following among electrical engineers.[105]

The subsequent convergence towards a single 60 cycle standard which had occurred by the turn of the century was made possible by there having been sufficient improvements in the 60 cycle rotary converter, so that its performance characteristics began to approach those of the 25 cycle converter. The final tipping of the balance however came about when it was found that the 60 cycle converter's performance could be further enhanced when used in conjunction with the high speed turbo-generators.[106] Beyond the turn of the century the

---

104 Lamme (1918), p. 70.

105 Lamme (1918), p. 83.

106 Lamme (1918), pp. 73-78.

only major exception to the 60 cycle system standard in America was that the Southern California system continued on 50 cycles as late as the 1950s. The prolonged existence of this anomoly has two explanations. First, as 50 cycles became the standard in many foreign countries, American manufacturing firms desiring to export to foreign markets did not feel unduly burdened by having to manufacture 50 cycle generation and end-use equipment.[107] Second, some electrical equipment, with no or only very slight alteration could be operated off either 50 or 60 cycle current which amerliorated the effect of there being two co-existing frequency standards.

A brief comparative glance at the contemporaneous situation in Germany lends support to the view that the convergence to a single a.c. frequency standard in the U.S. was indirectly delayed through the indirect influence of the gateway innovation (the rotary converter) that had lent greater momentum to the development of a.c.-based universal electricity supply systems. In Germany a standard frequency appears to have been arrived at rather more expeditiously than was the case in the U.S. But, there too, previous technological choices helped to determine the direction that would be pursued with regards to choosing a standard. Two factors influenced the German path. First, Emil Rathenau, long the head of AEG, the leading German electrical manufacturing firm, had always demonstrated an engineering preference for use of slower-speed, directly coupled generators from which high frequencies such as the 133.33 cycles, often initally used in a.c. lighting systems, could not be obtained. Hence the pressure for settling on a lower frequency. How low, however, would depend on the other equipment which would be used with the system. The answer that the standard should be a mid-range frequency--like the 50 cycle standard that actually was chosen--reflected the fact that Germany, unlike America, relied very little on the use of rotary converters in its supply systems.[108] This smaller reliance on rotary converters was in turn a consequence of the fact that less electricity was used for traction in Germany than in America. Statistics for the period 1910-12

---

107 Lamme (1918), p. 83.

108 Hughes (1983), pp. 128-9.

show traction responsible for only 31 percent of total kilowatt-hours sold in Berlin, whereas the same figure for Chicago was 69 percent.[109]

In summary, the evolution of a frequency standard in the U.S. and Germany had three notable features. First, in regards to the particular frequencies chosen, "accidents of history" clearly left an enduring mark. Frequencies were not arrived at on the basis of any clear perception of their long-run technical superiority; general acceptable ranges were determined by antecedent technologies and features of the current best-practice system design, but, the final outcomes were generally the product of a compromise. Second, a standard emerged with relative rapidly because one or two dominant manufacturers appear to have recognized its value in increasing the demand for equipment and in reducing its production costs. This source of commercial gain to the equipment suppliers derived from the gains which component compatibility afforded to utilities via power-sharing and load-balancing, and from the transferrability of electrical equipment across local utility boundaries in an era when physical networks were being extended, as well as from the advantages of being able to assemble larger batches of standardized equipment from interchangeable parts.

But, recognition by competing firms that establishing even a non-proprietory frequency standard would be desirable was not sufficient to insure a solution to the coordination problem, as evidenced by the experience of the first half of the decade of the 1890s in the U.S. In the event, the structure of the industry would appear to have had a key role in both the decisions of individual firms to choose a standard for their own production, and in determining whether or not industry-wide convergence on one standard proceded rapidly or stretched out over a long span of time. It should be emphasized that by the mid-1890s both the U.S. and German electrical manufacturing industries consisted of a few large firms. In the U.S. at least, one firm stood out as the initial leader in polyphase a.c. development--Westinghouse. The large size of the firms, and/or market leadership of one large firm made it possible for small historical events to tip the scales and bring about convergence on a standard once a sizeable part of the installed base of generating equipment had come from a single supplier, provided that supplier had adopted a uniform frequency standard.

---

109 Hughes (1983), p. 232.

Moreover, since most of the electrical engineers, other than those employed by utilities, worked for these few large manufacturing firms, it was relatively easy for engineers responsible for the design of equipment to stand united and firm against whatever tendencies there might be to subvert a standard in favor of higher performance, or meeting the specialized needs of individual customers. The oligopolistic structure of the market in the U.S., as well as that in Germany, may well have encouraged the formation of this professional stance on the part of electrical engineers, since it was tantamount to removing one dimension of non-price competition which would be difficult to monitor. The potential importance of internal policing of a commitment to uniform technical standards is evident, at least for the case of frequency standards, from the following recollection by Bejamin G. Lamme:[110]

> "Occasion has arisen, times without number, where an obvious solution of a given problem would lie in modification of the frequency to allow the use of apparatus and equipment already designed, but the engineers of the manufacturing organization have steadily held out against such a policy, regardless of the apparent need of the moment."

5. Concluding Observations

Charles S. Bradley, who filed for a U.S. patent on the rotary converter in 1888, would appear to deserve a larger place in accounts of the evolution of electric power networks than historians have accorded him. While Bradley's work has not been overlooked by the chroniclers of electrical inventions, the preceding discussion should serve to underscore several points about the emergence and impact of "gateway" technologies that, perhaps, have not been sufficiently appreciated--either by students of technological history or by applied microeconomists who have undertaken to analyze the strategic aspects of dynamic rivalries among increasing returns technologies.

First, network technologies are not static, and initial technical incompatibilities between variant formulations of such technologies, whether sponsored or unsponsored, can have their economic importance mitigated as a result of the ex post introduction of gateway innovations.

Second, the introduction of gateway innovations cannot be treated as exogenous, or autonomous developments in the evolution of a network

---

110 Lamme (1918), p. 85.

technology. Such innovations, and the inventions on which they are based, necessarily are conditioned upon an antecedent technological and economic context, and thus more likely than other classes of network components to qualify as having been "induced".

Third, although market rivalries among formulations of an emerging technology can be conveniently stereotyped for purposes of formal analysis as situations in which all decisions regarding the technical provision of compatibility between competing systems lie within the exclusive control of the respective contending system-sponsors, the games encountered by real-world players are not so manageable. Sponsors of actual or latent systems have to recognize that when the positive network externalities are large (which is to say, when the social costs of incompatibility and the potentially capturable rents from its elimination are large) it is more likely that there will be private, market-induced responses from "third parties" . In the historical case examined here, the profit-potentialities inherent in creating a hyrid system that would integrate elements from the d.c. and a.c. electrical supply technologies, each of which had its special area of advantage, elicited the development of a converter device by an independent inventor-entrepreneur who was not allied with either of the main system sponsors, Edison or Westinghouse.

Fourth, once there has been a "third party" induced intervention of this kind, supplying a gateway technology, existing system-sponsors have less to gain from trying to further develop their proprietary technology so as to make it the basis of a more comprehensive network that not only mimics, but, improves upon the features of the "patched" or "hybrid" system employing converters. An enterprise such as Edison's, whose comparative advantage in research leading to patentable innovations forming the main constituents of a whole new technology from which they could extract large rents, is likely to see it as desirable to exit a maturing technical field at such a point. This would tend to transfer the commercial exploitation of the original system concept into the hands of firms that had an absolute advantage in its production and marketing, rather than in pursuing more radical design improvements. In this way, the advent of more elaborate "gateway" innovations may directly contribute to--as well as being indicative of--the retarded pace of technical advance accompanying the process of industrial "maturation".

Fifth, once a gateway innovation appears on the scene, even though a particular class of devices may be symmetric in their ability to convert--as in our example, from d.c. to a.c. and back--the impact of its introduction upon the competition between contending variants need not be symmetric; it may disturb a delicate "balance" of market advantages based on the heterogeneity of specialized user-needs, and so assist one particular variant to emerge as the standard for the enlarged network. In the case at hand, it has been seen that the system which happened to hold the cost advantage in electricity transmission at the point in time at which the rotary converter was developed was favored to become the main supply current.

Sixth, due to the strong elements of interrelatedness that are prevalent in network technologies, when a gateway innovations is introduced within an already complex system, its reprecussions upon the development of other technical components are likely to be quite far-reaching, but do not necessarily promote the elimination of technical incompatibilities throughout the system. In the instance examined, it was seen that rotary converter was implicated directly in the initial establishment of a dual standard for a.c. frequencies--at 25 cycles and 60 cycles per second--rather than a uniform single standard. It had this consequence because, in extending the scope of a.c. networks to a wider and more heterogeneous population of electricity users, a conflict was created between the frequency requirements of the electric lighting market and the frequency needs of customers who would use converters to supply power in the form of direct current for electric trams and railways.

Seventh, gateway technologies tend to have their own life cycle, inasmuch as once they have formed an economically feasible bridge across some technical incompatiblity, the passage of traffic over that particular bridge is increased with increasing system usage, thereby directing attention to the cost-savings that could be achieved if it were entirely eliminated. Once the rotary converter had tipped the balance towards the growth of a.c.-based electricity supply networks in the U.S., the subsequent inducements to eliminate the fixed costs of converter equipment, as well as the costs of energy losses entailed in the conversion itself, acted to focus research and development efforts on the part of electrical manufacturers in the direction of designing end-use equipment, appliances, and motors, immediately compatible with the dominant form of (alternating) current. Hence, over time, cost and performance characteristics of a.c. motors tended

to improve relative to those of d.c. motors, until in the U.S. only users with very specialized needs--such as high torque at low rpm, for traction work and heavy drilling, or minimal electromagnetic interference as in electronic audio equipment--continued to employ converters to drive direct current appliances.

Eighth, through the effects they have in altering the composition of heterogeneous customers being served by a network, gateway technologies can have indirect consequences for the distribution of eventual network benefits; they may even have effects that prove disadvantageous to the class of customers originally served by a more restricted network. Had polyphase a.c. emerged as the electricity network standard less precipitously than it did in the U.S., the market for d.c. motors and the opportunies offered for the realization of static and dynamic economies of scale in their production would have been enlarged. Under competitive conditions in the supplying industry, this would have been particularly beneficial to traction companies and other specialized users of d.c. motors, but rather less so for those industrial power users, such as cotton textile mill operators, who would particularly value the "non-sparking" feature offered by the a.c. motor. Thus, while there is no intention here of suggesting any conclusion be drawn regarding the global social optimality of these complex and contingent developments, or other aspects of the outcomes affected by the way in which the "contest of the currents" was resolved, neither should we dismiss from consideration the redistributional effects which these remote historical events surrounding the induced advent of a gateway innovation turned out to have upon the various classes of customers who eventually came to be served by the electricity supply industry.

Enough has been said throughout the preceding discussion--particularly in regard to the influence that non-technological decision contexts, and events such as political legislation, business mergers and exits effecting market structures--to make it plain that the examination of the role of gateway technology innovations can exposes only one narrow facet of the complex process through which network industries evolve. But it is a facet worth defining, and further polishing. Though usually relegated to the category of "minor" innovations, the class of gateway technologies and their economic consequences clearly deserve closer empirical study and more systematic theoretical analysis than they have yet received.

Acknowledgments

The authors are, respectively, the William Robertson Coe Professor of American Economic History, and a Ph.D. Candidate in the Department of Economics at Stanford University. We are indebted to a number of people for their comments and suggestions in response to earlier drafts and seminar presentations of some of the material contained in this paper, especially: Edward Constant, Shane Greenstein, Marlon Gunderson, Jonathan R. T. Hughes, Thomas Parke Hughes, W. Edward Steinmueller and Peter Temin. This research received financial support through the High Technology Impact Program of the Center for Economic Policy Research at Stanford University.

References

Alvarado, F. L. and F. H. Lansetter [1984], "Methodology for Analysis of Multi-Terminal HVDC Systems," in Clark W. Buyllard and P. J. Womelodoff, eds,. Trends in Electric Utility Research, New York: Pergamon Press. pp. 210-30.

Arthur, W. Brian [1984], "Why a Silicon Valley?: The Pure Theory of Locational Choice," Technological Innovation Project Workshop paper, Department of Economics, Stanford University, November.

Arthur, W. Brian [1985], "Competing Technologies and Lock-in by Historical small Events: The Dynamics of Allocation Under Increasing Returns," Technological Innovation Project Working Paper No. 4, Center for Economic Policy Research (Publication No. 43), Stanford University.

Arthur, W. Brian, Yuri M. Ermoliev, and Yuri M. Kaniovski [1985], "Strong Laws for a class of Path-dependent Urn Processes," in Proceedings of the International Conference on Stochastic Optimization, Kiev, 1984, Munich: Springer-Verlag.

Barstow, W. S. [1901], "Notes on the Alternating Current System of Distribution," AIEE Transactions, Volume XVIII, pp. 849-53.

Berg, Sanford V. [1985], "Duopoly Compatibility Standards with Partial Cooperation and Stackleberg Leadership," Working Paper, University of Florida Department of Economics. August.

Besen, Stanley M. and Leland L. Johnson [1986], Compatibility Standards, Competition, and Innovation in the Broadcasting Industry, Santa Monica, Ca.:The RAND Corporation. November.

Bresnahan, T.R. and P.A. David [1986], "Strategic Effects in the Diffusion ofInnovations: Testable Implication and Preliminary Findings for the Case of Automatic Teller Machines in U.S. Banking Markets," a paper presented to the Conference on Innovation Diffusion, held at the University of Venice, Italy, March 17-22, 1986. (Technical Papers Series of the High Technology Impact Program, Center for Economic Policy Research at Stanford University, Stanford, CA.)

Brock, Gerald [1975], "Competition, Standards and Self-Regulation in the Computer Industry," in R. Caves & M. Roberts (eds.), Regulating the Product: Quality and Variety, Cambridge, MA: Ballinger Publishing Company.

Bunn, Julie A. [1987], "Politics, Markets, and Engineers: A Re-examination of the "Engineering Backwardness" and "Engineering Culture" Hypotheses as Explanations for Lags in British Electrical Industry Development," Paper prepared for the Social Science History Seminar, Stanford University, Autumn 1987.

Byatt, I.C.R. [1979], The British Electrical Industry, 1875 - 1914: The Economic Returns to a New Technology. Oxford: Clarendon Press.

Carlton, Dennis W. and J. Mark Klamer [1983], "The Need for Coordination Among Firms, with Special Reference to Network Industries," in The University of Chicago Law Review 50:446-465.

Cheney, Margaret [1981], Tesla: Man Out of Time. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Cusumono, Michael [1985], "Note on VTR Industry and Market Development: Japan, the U.S., and Europe, ca. 1975-1985," Harvard Business School.

David, Paul A. [1975], Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century. New York: Cambridge University Press.

David, Paul A. [1985], "Clio and the Economics of QWERTY," American Economic Review 75, 2 (May): 332-337.

David, Paul A. [1986a], "Understanding the Economics of QWERTY: The Necessity of History," in Economics History and the Modern Economist, W.N. Parker, ed., Oxford: Basil Blackwell.

David, Paul A. [1986b], "Some New Standards for the Economics of Standardization in the Information Age," Technological Innovation Project, Working Paper No. 11, Center for Economic Policy Research, Stanford University. August. (forthcoming in The Economics of Technology Policy, P. Dasgupta and P. L. Stoneman, eds., Cambridge University Press.)

David, Paul A. [1987], "The Hero and the Herd in Technological History: Reflections on Thomas Edison and the *Battle of the Systems'", Center for Economic Policy Research Publication No. 100, Stanford University. July.

David, Paul A. , with Julie Ann Bunn [1987], "'The Battle of the Systems' and the Evolutionary Dynamics of Network Technology Rivalries," High Technology Impact Program Working Paper No. 15, Center for Economic Policy Research, Stanford University. January.

David, Paul A. , and Trond E. Olsen, [1986] , "The Equilibrium Dynamics of Diffusion when Incremental Technological Innovations are Foreseen," Ricerche Economiche (Special Supplement on Innovation and Diffusion) 1986, n. 4 .

Electrical World (articles without identified authors):

[1887], "Edison vs. Westinghouse," January 1, pp. 7-8.

[1887], "Incandescent Lights on High Tension Circuits," February 26, p. 110.

[1887], "The Distribution of Electricity by Secondary Generators,"

Hughes, Thomas P. [1962], "British Electrical Lag: 1882-1888," Technology and Culture, 3(1), Winter: pp. 27-42.

Hughes, Thomas P. [1983], Networks of Power: Electrification in Western Society, 1880-1930. Baltimore: The Johns Hopkins University Press.

Jarvis, C. Mackenchnie [1967a], "The Generation of Electricity," Ch. 9 in Charles Singer, E.J. Holmyard, A.R. Hall, and Trevor I. Williams, editors, A History of Technology, Volume V:Late 19th Century, 1850-1900. Oxford: The Clarendon Press, pp. 177-207.

Jarvis, C. Mackenchnie [1967b], "The Distribution and Utilization of Electricity, " Chapter Ten in Charles Singer, E.J. Holmyard, A.R. Hall, and Trevor I. Williams, editors, A History of Technology, Volume V: Late 19th Century, 1850-1900. Oxford: The Clarendon Press, pp. 208-234.

Josephson, Mattew [1959], Edison: A Biography. New York: McGraw-Hill.

Jones, Payson [1940], A Power History of the Consolidated Edison System. New York: New York Consolidated Edison.

Katz, Michael L. [1986],"The Economics of Standarization in Network Industries." Paper Presented to the Fourteenth Annual Telecommunications Policy Research Conference, Airlie House, Va.

Katz, Michael L. and Carl Shapiro [1985a], "Network Externalities, competition and compatibility," American Economic Review 75 (May): 424-440.

Katz, Michael L. and Carl Shapiro [1985b], "Technology Adoption in the Presence of Network Externalities," Discussion Paper in Economics No. 96, Woodrow Wilson School, Princeton University.

Kindelberger, Charles P. [1983], "Standards as Public, Collective, and Private Goods," Kyklos 36 (Fasc. 3):377-96.

Kline, Ronald [1987], "Science and Engineering Theory on the Invention and Development of the Induction Motor, 1880-1900," Technology and Culture, 28(2), July, pp. 283-313.

Lamme, B. G. [1918]. "The Technical Story of the Frequencies," A.I.E.E. Transactions Volume XXXVII, Part I.

Lamme, Benjamin G.[1926], Benjamin Garver Lamme, Electrical Engineer: An Autobiography. New York: G.P. Putnam's Sons.

Lineff, W. [1888], "Accumulators versus Direct Currents for Electric Traction," Electrical World, April 14.

Leupp, Francis E. [1919]. George Westinghouse: His Life and Achievements. London: John Murray.

March 26, pp. 156-8.

[1887], "Report on the Committee on Electrical Distribution by Alternating Currents," August 20, pp. 92-3.

[1888], "Alternating Current or Storage Battery, " January 21, p.25.

[1888], "Alternating vs Continuous Current," March 31, p. 159.

[1888], "The Comparative Danger of Alternating vs. Direct Current," September 8, p. 126.

[1889], "Mr. Edison on the Dangers of Electric Lighting," November 2, pp. 292-3.

The Encyclopedia Britainnica, A Dictionary of Arts, Literature and General Information [1910-1911], Eleventh Edition. 29 vols. New York: The Encyclopedia Britannica Co.

Evans, G.T. [1892], "Some Phases of the Alternating Current," Electrical World, January.

Farrell, Joseph and Garth Saloner [1985a], "Standardization, Compatibility and Innovation," Rand Journal 16 (spring): 70-83.

Farrell, Joseph and Garth Saloner [1985b], "Economic Issues in Standardization," Working Paper No. 393, Department of Economics, Massachusetts Institute of Technology, October.

Farrell, Joseph and Garth Saloner [1985c], "Installed Base and Compatibility, with Implications for Product Preannouncements," Working Paper No. 385, Department of Economics, M.I.T., August.

Farrell, Joseph and Garth Saloner [1986], "Competition, Compatibility, and Standards: The Economics of Horses, Penguins and Lemmings," forthcoming in Product Standardization as a Tool of Competitive Strategy, ed. H. Landis Gabel. Amsterdam: North-Holland.

Fleming, J.A. [1921], Fifty Years of Electricity: The Memories of an Electrical Engineer. London: Published from the offices of Wireless World.

Fudenberg, T. and J. Tirole [1985], Dynamic Models of Oligopoly, (Fundamentals of Pure and Applied Economics, Volume 3: Theory of the Firm and Industrial Organization Section) London: Harwood Academic Publishers.

Hemenway, David [1975], Industrywide Voluntary Product Standards, Cambridge, Mass.: Ballinger Press.

Hughes, Jonathan [1986]. The Vital Few. New York: Oxford University Press.

McDonald, Forrest [1962], Insull. Chicago: University of Chicago Press.

National Electrical Manufacturers Association [1946], A Chronological History of Electrical Development. New York: NEMA.

Parsons, R.H. [1940], The Early Days of the Power Station Industry. Cambridge: University Press.

Passer, Harold C. [1972], The Electrical Manufacturers: 1875-1900: A Study in Competition, Entrepreneurship, Technical Change, and Economic Growth. New York: Arno Press.

Pfrannkuche, A. [1887], "Long Distance Distribution of Electric Energy," Electrical World , April 2.

Prout, Henry G. [1921]. A Life of George Westinghouse. New York: The American Society of Mechanical Engineers.

Puffert, Douglas J. [1986], "Standardization of Railway Gauge in Britain and America," Social Science History Workshop Paper, Department of Economics, Stanford University, May.

Rosenberg, Nathan [1969], "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices," Economic Development and Cultural Change, Vol.18, No.1 (October).

Rosenbloom, Richard [1985], "Managing Technology for the Longer Term," in K. Clark et al., eds., The Uneasy Alliance, New York: Praeger.

Scott, Charles F. [1901], "Alternating Current as A Factor in General Distribution for Light and Power," AIEE Transactions, Vol. XVIII, pp. 843-48.

Sharlin, Harold I. [1963], The Making of the Electrical Age. New York: Abelard-Schuman.

Stanley, W. [1912], "Alternating Current Development in America," Franklin Institute Journal 68: 568-73.

Stillwell, Lewis B. [1934], "Alternating Current Versus Direct Current," Electrical Engineering. 53(5), pp. 708-10.

Sumpner, W.E. [1890], "Some Peculiarities of Alternate Currents," Electrical World, January.

Weeks, Walter [1981]. Transmission and Distribution of Electrical Energy. New York: Harper & Row, Publishers.

Woodbury, David O. [1960], Elihu Thomson: Beloved Scientist, 1853-1937.Boston: The Museum of Science (1944,1960).

Zorpette, Glen [1985], "HVDC: Wheeling Lots of Power." IEEE Spectrum, June, pp.30-36.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Coordination Through Committees and Markets

by

Joseph Farrell* and Garth Saloner**

April 1987

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

*Department of Economics, University of California, Berkeley.

**Department of Economics, M.I.T., and National Fellow, Hoover Institution, Stanford University. Saloner would like to thank the National Science Foundation (grant IST-8510162) and the Hoover Institution for financial support.

We are grateful to Ken Hendricks and Oliver Hart for helpful comments.

IR013166

ABSTRACT

We discuss two common mechanisms for achieving coordination, with particular reference to the choice of compatibility standards. The first involves explicit communication and negotiation before irrevocable choices are made: it represents what standardization committees do. The second mechanism, by contrast, involves no explicit communication, and depends on early irrevocable choices: the hope is that one agent will choose first and the other(s) can then follow. This is a simple version of "market leadership." We analyze these two mechanisms in a simple model, and show that the committee is more likely to achieve coordination. Although it is slower, it outperforms the market mechanism, even when we allow for the value of speed. We also examine unilateral preemptive actions that subvert the committee system and show that, far from marring its performance, they actually improve it.

1. Introduction

More than a hundred thousand people meet regularly in committees with the goal of reaching agreement on product and interface compatibility standards. The resources devoted to this formal standardization activity have roughly doubled in the last decade. Yet there is disagreement over whether these "standardization committees" are a good way to organize attempts at coordination. In this paper, we introduce a theory of how such committees work, and compare their performance with that of a market left to itself.

Committees set standards in a wide range of industries, from lumber to local-area-networks. The American National Standards Institute (ANSI), a private body, has about a thousand corporate members and several hundred affiliated professional organizations. It collects, disseminates and coordinates standards set by member organizations and by industry groups. ANSI is also the United States' representative at the International Organization for Standardization (ISO), which works to harmonize standards internationally and to set standards when a national approach seems inadequate. The ISO, for instance, is responsible for the Open Systems Interconnection (OSI) reference model in mainframe computers, which provides a framework for achieving compatibility.

The computer and telecommunications industries, with their rapid growth, urgent demands for compatibility, rapid innovation of new products and services, and increasing

fragmentation, have driven much of the recent growth in formal standardization. Consider the following three examples: First, the CCITT has undertaken an ambitious project, Integrated Services Digital Network (ISDN), to set standards for the rapidly progressing merger of the two industries.[1] ISDN alone has dozens of active committees working on a wide range of problems. Second, the Federal Communications Commission (FCC) recently nominated the Exchange Carriers Standards Association, a private organization formed after the Bell breakup, to try to negotiate standards that will give competitors of the Bell Operating Companies "comparably efficient interconnection" to enable them to compete on an equal footing. Finally, in the computer industry itself, the recently-formed Corporation for Open Systems, funded by over thirty of the largest information-processing companies in the nation, aims "..to provide an international vehicle for accelerating the introduction of interoperable multivendor products and services operating under agreed-to open systems interconnection... to assure (sic) acceptance of an open network architecture in world markets".[2]

Such careful and explicit cooperation is a natural response to the need for coordination. Yet economists have almost completely ignored this activity, and have focused instead on the coordination that may be achieved in the marketplace

---

[1] See Besen and Saloner (1987) for a detailed discussion of ISDN standard-setting.

[2] Quote from Rutkowski (1986). See Besen and Saloner (1987) for more detail on these institutions and for further references.

without explicit collaboration. The price system, often extolled for its coordinating abilities, is inadequate to this task, however, since the problem is not how much to produce or how, but what to produce, and prices do not exist until that decision has been taken.

Instead, the market mechanism that we examine works by making product choice sequential instead of simultaneous. If one important agent makes a unilateral public commitment to one standard, others then know that if they follow that lead, they will be compatible at least with the first mover, and plausibly also with later movers. This bandwagon mechanism can sometimes achieve rapid and effective coordination[3]. A recent example is the protocol for scrambling satellite signals to cable-TV stations. Several systems were available, but after Home Box Office, the largest player in the market, chose VideoCipher, other users rapidly followed suit, and coordination was thus achieved[4]. A similar pattern was typical in US telecommunications before the Bell breakup: Bell made its choices, and the smaller players such as GTE naturally followed.[5]

---

[3] See Farrell and Saloner (1985) for a treatment of how bandwagons can solve certain coordination problems, notably that of coordinating adoption of a new and superior standard when an inferior one has become established. Farrell and Saloner (1985, 1986b) also discuss some limitations of the bandwagon process other than those described in the present paper.

[4] See Besen and Johnson (1986).

[5] See e.g. Rutkowski (1986) and Sullivan and Zader (1985).

The market mechanism does not always achieve standardization, however. For example, in AM stereo broadcasting several incompatible systems are in use and overall adoption has been slow.[6] In VCRs neither Sony nor Phillips felt obliged to follow the other, and the incompatible systems they marketed have persisted in the market. The worldwide adoption of color television was similar: three main incompatible systems (NTSC, PAL, and SECAM) competed, and no adopter had enough power to start a truly irresistible bandwagon[7]. These examples suggest that the "bandwagon" mechanism for coordinating is imperfect when there is no predetermined leader and when people have different preferences across standards. This result has also emerged from the growing theoretical literature on standardization[8].

Formal negotiation through a committee can avoid some of the problems that mar the performance of the market mechanism. In particular, explicit communication and the commitment not to adopt without agreement prevent the problem of incompatible adoption that often arises in a market where two or more firms try to lead the bandwagon. But these committees too are imperfect coordinators. Often, by the time a committee is convened, participants have vested interests in incompatible

---

[6] See Besen and Johnson (1986).

[7] See Crane (1979).

[8] See for instance Farrell and Saloner (1985, 1986b), Katz and Shapiro (1985, 1986a,b), and Matutes and Regibeau (1986). See Besen and Saloner (1987) for an overview of this work and for a framework of the determinants of the standard-setting process.

positions.[9] Thus, in addition to the committee's difficult tasks of defining its problem, formulating alternatives, learning their relative merits, and coordinating with other committees, it must resolve conflict. Since the "consensus principle", which is widely accepted in voluntary standard setting, rules out a simple vote and requires committees to seek a more compelling consensus (though not necessarily strict unanimity), this can lead to a battle of wills in committee, while the market waits.[10] Such waiting is costly, whether simply because of delay or because eventually the participants can wait no longer and the chance for coordination has been missed.

Both the market (bandwagon) and the formal (committee) system, then, are imperfect. Which is better? What does that depend on? What are the advantages and disadvantages of each? These questions are not only of academic interest: practitioners often seem uncertain about the relative merits of the two approaches. For example, the FCC has sometimes suggested that more standardization decisions should be "left to the market,"[11] - a view echoed in the remarks of Dr Theodor Irmer, Director of

---

[9] This has occurred, for example, in modems, OSI, CD-ROM, railroads, and personal computers.

[10] See Verman (1973) and Sanders (1972) for a discussion of the consensus principle.

[11] For example, this occurred in the case of AM stereo and to a lesser extent in TV stereo.

the CCITT, at a recent conference[12]. Yet Irmer himself is widely regarded[13] as the driving force behind ISDN, the epitome of formal standardization; and, as mentioned above, the FCC urges committee standardization of an "open network architecture". These apparent contradictions pose a major public-policy problem: what, if anything, should government do to help the standardization process? What advice can economists give to participants? In particular, can we say anything useful about the relative merits of the different ways to organize standardization? We hope that our theory will help us and others to think about these problems.

1.1 Methodology

We study the case where everyone would prefer any proposed coordinated (standardized) outcome to the result of each going his own way[14], but where they disagree on which of the coordinated outcomes is better. Such conflict is commonplace for many reasons. Sometimes it arises because different participants

---

12 "Product Standardization as a Competitive Strategy," INSEAD, June 9-10, 1986. Proceedings forthcoming (North-Holland), edited by H. Landis Gabel.

13 See e.g. Rutkowski (1985), p.9.

14 Even where the benefits to compatibility are large, this is not inevitable. Katz and Shapiro (1986b) show how firms' incentives to choose compatibility or incompatibility are affected by the dynamics of competition between them. And even if compatibility is preferable, the disadvantage of having to catch up on a rival's proprietary standard may be so large as to overcome that preference, so that compatibility may be impossible to reach unless a compromise is possible. For example, the European computer group X/OPEN chose UNIX as its standard operating system because no member was prepared to switch to another's previously proprietary standard.

plan to serve different segments of the market. For example, in videotex, some users would want a graphics capability such as that provided by NAPLPS[15], which some systems provide; but other users would prefer the lower cost and greater speed of an ASCII system. In other cases, there may be no important conflict among users, but one vendor has a competitive advantage in a particular system, most likely one that it developed and in which it may retain proprietary rights, or at the least an advantage of experience. For example, in AM stereo, users are essentially indifferent among the competing systems, but each vendor had a vested interest in its own system that prevented agreement on a standard. Alternatively, there may be legitimately different opinions about which standard would be best for the entire industry in the long run. Finally, at the international level, nations may hope to protect domestic firms by pressing for certain standards.[16] In any case, we suppose that there is conflict: if not, then agreement will likely be rapid and easy under almost any institutional structure.

In such a problem, where each agent prefers a different coordinated outcome, each agent's best strategy depends on that chosen by the other(s). If your rival insists on his proposal,

---

[15] North American Presentation Level Protocol Syntax. See Besen and Johnson (1986) p. 80.

[16] See Crane (1979) on the battle over color TV standards. A very similar fight is proceeding on the new high-definition television (HDTV) front. European, especially French, government representatives at the CCIR's plenary meeting in 1986 blocked the working committee's proposal to adopt NHK's system, which American and Japanese interests favored.

then it is wise to yield; but perhaps he is only bluffing, and will yield if you insist on yours. Such strategic choices are the subject-matter of game theory. Indeed the problem has the payoff structure of the classic "battle-of-the-sexes".[17] Game theory has been slow to model the problem of coordination, however, because Nash equilibrium assumes it away: it is an equilibrium for us both to choose my preferred standard, and another for us to choose yours.

These pure strategy Nash equilibria, however, are quite unrealistic without a convincing story of the mechanics of coordination. It is absurd to claim, by choice of solution concept, that either I will know to yield while you hold firm or vice versa, and that there will be no confusion, no bluffs, and no accidents. At the least, we can demand a model of the process or institutions that leads to such remarkable coordination. While it is perhaps legitimate to assume, for the purpose of analyzing other phenomena, that some such process might exist in the background, it would certainly be unreasonable to do so when we are setting out precisely to study the mechanics of coordination.

To model institutions that promote coordination, therefore, we must use a solution that allows for some failure of coordination. Such a failure might be expected to arise if there

---

[17] In this classic game of coordination, a man and a woman are choosing between the ballet and a prize fight. The main concern of each is to be in the company of the other (regardless of where they go), however he prefers the prize fight and she the ballet.

is no pre-play negotiation and if the players are symmetric so that (at least ex ante) they can be expected to behave in the same way. An obvious way to model such symmetric behavior is by concentrating on symmetric equilibria. The symmetric equilibrium of the battle-of-the-sexes is a mixed strategy equilibrium. Each player sometimes insists on his/her own way, and sometimes yields.

We model three processes of coordination and compare their relative speed, efficacy, and payoffs. First, we consider what happens when the players can meet (many times) in order to agree on a joint action. Modeling the committee in this way produces a many-period game, which, it turns out, itself has the basic structure of a "battle-of-the-sexes" game each period. Analyzing the symmetric (mixed-strategy) equilibrium of that game yields a description of the effectiveness of committee negotiation. The overall game is analogous to a "war of attrition".[18]

The second coordination process that we model is that of pure action: the market or bandwagon process. Our model of that process produces another many-period game, which again has the basic structure of the battle-of-the-sexes each period. The analysis of its symmetric equilibrium provides conclusions about the nature of the bandwagon process. The overall game is a

---

[18] See Hendricks and Wilson (1986), Fudenberg and Tirole (1985).

"grab-the-dollar" game.[19]

The last coordination process that we consider is a "hybrid" in which players use both bandwagon and committee strategies. In any period, as well as negotiating in committee, an agent can also commit to a course of action in the market. The reason for considering a hybrid is that if all other agents are talking and not adopting, then any one agent has a clear chance to seize the lead by unexpectedly adopting. If players cannot commit themselves not to do that, then it is only the fear that another agent will simultaneously try to do the same that prevents any one agent from adopting. Therefore a "pure" committee system in which no one acts until agreement is reached is an equilibrium only if the players can commit themselves to it.

1.2 Summary of the Model and Results

We consider a model with two possible but mutually incompatible new technologies or systems: A and B. There are two agents (firms or users), whose preferences are as follows: it is common knowledge[20] that agent 1 prefers A, and agent 2 prefers B; but each would prefer to adopt the less preferred option than be incompatible. Thus, agent 1 is best-off if both he and agent 2 choose system A; his next preference is for them both to choose B; next is the outcome where each chooses his own (1 chooses A, 2 chooses B); the worst is if 1 chooses B and 2 chooses A.

---

[19] This game is discussed, for example, in Fudenberg and Tirole (1985). It was first proposed by Richard Gilbert.

[20] A simple straw vote, which we need not even model, would certainly establish if the opposite were true.

We begin by comparing the pure committee and bandwagon systems. In Section 2, we consider the case in which there is a fixed date by which decisions must be made, but nothing is gained by making decisions earlier; activity before that date (the "deadline") is solely for the purpose of coordination. This allows us to analyze the "reliability" of the mechanisms without worrying about their speed. Simple calculations would suffice to compare the outcomes with one or two periods before the deadline, but we wish to extend the model beyond that. To do so, we develop a new analytical tool. We reduce each many-period game, with its plethora of (normal-form) strategies, to a simple 2-by-2 game in which some of the payoffs are endogenous. A simple game-theoretical result then enables us to compare equilibrium payoffs.

We find that the committee unambiguously outperforms the bandwagon system, but that the difference in payoff vanishes in the limit as the number of periods goes to infinity. Although the payoffs become equal in the limit, the behavior of the systems does not. In a committee, with many periods to go, nothing is likely to happen for a long time: almost certainly, neither player will yield until the deadline approaches. In the bandwagon game, by contrast, early action is likely: in fact, there is always a finite expected time to a definitive outcome, and with many periods a decision is essentially certain before the deadline.

The bandwagon outcome, however, even if reached early, is not necessarily good: often, simultaneous and incompatible adoption mars the bandwagon's coordination. The committee is less likely to reach an outcome before the deadline, but if it does so then the outcome is certainly coordinated. If the deadline arrives with no committee decision, then coordination is still possible by chance, but so are the other two outcomes, one of which (player 1 adopts B, and 2 adopts A) is worse than anything that can happen in the bandwagon game.

In Section 2, we show that the committee does better than the market when there is no value attached to speed. However, the committee is slower. One might naturally expect, therefore, that allowing for the importance of speed might reverse this conclusion. And speed is often important: it is often the case that an earlier decision is better for society and for the agents themselves.[21] We represent this in Section 3 by discounting the payoffs by a constant factor for every period in which no coordination is attained. An analytical technique similar to that used in Section 2 works in this case also, and we find again that the committee system gives unambiguously better payoffs in equilibrium.

In Section 4, we consider the hybrid system. Since this game is in some sense a mixture of those in the pure bandwagon and committee games, we might expect an intermediate payoff. In

---

[21] This is not always so. Since we are considering nascent technologies, new information may arrive, and delay may be valuable.

fact, however, we show that the hybrid game gives strictly greater payoffs even than the "pure" committee system! Intuitively, the reason is that if neither player starts a bandwagon, they have a chance to talk (and thus another chance to coordinate) before the next stage of the game; similarly, if they fail to reach agreement in the committee room, there is a chance of coordination in the marketplace. Thus they have "two chances" to coordinate each period. One might feel that this is "unfair:" that we should instead model the hybrid as (perhaps) alternating bandwagon-style and committee-style moves. But we believe that the hybrid's vigor revealed in this result is genuine: if the two processes do not destructively interfere with one another (as our model suggests they do not), then it is in fact desirable to have both simultaneously in action. This result suggests that, while committees are better than the pure bandwagon system, they are even better if subverted by preemptive action outside the committee!

2. The Model

2.1 Assumptions and Description

We denote the value to each player of successful coordination by c, and the value of being on his own privately preferred system as a. The case of interest is where $c > a > 0$, since otherwise each simply adopts his own preferred alternative. These preferences can be represented in the

following payoff matrix, a modified version of the battle-of-the-sexes:[22]

		Player 2 A	Player 2 B
Player 2	A	a+c, c	a, a
	B	0, 0	c, a+c

Figure 1: The Final Game

We begin by describing the mixed-strategy equilibrium of this game, which we call the "final game," since it must be played if no agreement is reached before the deadline. To make player 1 indifferent between his first and second actions, as must be the case in a mixed-strategy equilibrium, we require that $p(a + c) + [1 - p]a = p.0 + [1 - p]c$, where p is the probability that player 2 chooses system A. Hence $p=(c-a)/2c$. Since $a < c$, p is a legitimate probability. Moreover, $p < 1/2$: each player chooses his private preference more than half of the time in the mixed-strategy equilibrium.

The value of this mixed-strategy equilibrium to each player is then easily calculated as:

$$V = (c + a)/2 < c.$$

We think of this value as representing the players' payoffs from failing to agree before they must actually choose. Notice that V is actually less than each players less preferred pure-strategy equilibrium outcome, so that either player would in fact

---

[22] The minor modification here is that the uncoordinated outcomes have unequal payoffs.

prefer to give in for sure rather than receive V. However, in equilibrium, each holds out to some extent for his better payoff. While this outcome, V, may seem an unduly pessimistic forecast, that is probably because, in our experience, it is rare for there to be no coordinating mechanism at all. Since we will use V only to represent the outcome if the coordination mechanism in use fails completely, we find it entirely reasonable.[23]

2.2 Committee Game with n Periods

Now suppose that, before the game of Figure 1 must be played, the players have n chances to reach verbal agreement. That is, each can (simultaneously) announce "insist" or "concede"; if just one insists then he gets his way (this is an equilibrium of the subgame, and seems the plausible one), while if both insist or if neither does then they meet again the following period. If no agreement has been reached after n meetings then they play the mixed-strategy equilibrium of Figure 1.

To analyze this game, we could write out a very large payoff matrix for the normal-form game: each player has two choices at each date, which he can condition on history. However, we can describe the players' behavior in the first period in terms of a two-by-two matrix, if we use an endogenous "continuation payoff" in that matrix. In particular, if both

---

[23] Another apparently appealing representation of the outcome of no coordination would be that each player adopt his preferred system: 1 adopts A and 2 adopts B. But this would not be an equilibrium: either player unilaterally would prefer to change his choice.

"insist," so that no agreement is reached, they get V(n-1), the value of having (n-1) meetings remaining. If both "concede," it is less clear what we should assume. For the moment, we will be a little harsh on the players and suppose that they get V(n-1) in that case too. (We show below that our results are only strengthened if we make a more generous assumption.) Therefore the "reduced-form" payoff matrix is:

		Player 2	
		Concede	Insist
Player 1	Insist	a+c, c	V(n-1), V(n-1)
	Concede	V(n-1), V(n-1)	c, a+c

Figure 2: Committee Game with n Periods Remaining

If V(n-1) < c, as is certainly the case when n = 1 (V(0) is V derived above), then this game has the same general "coordination-game" form as the original (Figure 1). That is, the diagonals are better for both players than the off-diagonals, but they differ on which of the two diagonals is better. The value of the mixed-strategy equilibrium of this game is:

$$V(n) = p(n)(a + c) + [1 - p(n)]V(n - 1)$$
$$= p(n)V(n - 1) + [1 - p(n)]c. \qquad (1)$$

Since by inductive hypothesis V(n - 1) < c, it follows that V(n - 1) < V(n) < c. Since we have an increasing sequence bounded above, we must have convergence, say to V^*. It is also easy to see from (1) that p(n) decreases in n, and so also has a limit, p^*. Therefore,

$$V^* = p^*(a + c) + [1 - p^*]V^* = p^*V^* + [1 - p^*]c,$$

whence (since we know V^* cannot be $(a + c)$),

$$p^* = 0,$$

$$V^* = c.$$

In other words, when there are many periods of negotiation, there is very little concession early on ($p^* = 0$). The cumulative effect of the negotiation is to give each player almost as good a payoff as if they agreed immediately on the system he prefers less. Only the compatibility benefits are achieved, on average; the benefits on which they disagree are dissipated in imperfect coordination. Obviously, when a is small, this is quite a good outcome; when a is close to c, each gets c instead of the two together getting almost $(c + c) + c$, so about a third of possible payoff is wasted by their inability to coordinate more effectively.[24]

It is also of interest to examine the overall probabilities of specific outcomes. Let x^* be the probability of agreement, Then the payoff to the two players together is

---

[24] Notice that the committee could do better by simply commiting to toss a coin to choose between A and B. However, such a commitment would not be credible since the losing player would have an obvious incentive to renege on the commitment and call for a new coin toss. Of course it might be argued that the same kind of commitment is required in the committee game since, there too, the losing player would always want to change his mind ex post. However, such a change of mind is not costless in the committee game (as it is in the case of a coin toss), since any new decision can only be made when the committee reconvenes the next period. In fact it follows from the fact that $V(n-1) < c$, that such a defection is strictly unprofitable: a player who concedes is grateful that coordination has been achieved.

$x^*(2c + a) + (1 - x^*)V$, which must be equal to 2c. Hence (substituting for V), $x^* = (3c - a)/(3c + a)$.

Thus if a is small there is almost certain to be agreement in advance; if a is almost equal to c there is about a one-half probability of agreement in advance. If there is no agreement in advance, there is of course one last chance, as indicated in the analysis of the final game above. The overall probability of coordination is therefore

$$x^* + (1 - x^*)2p(1 - p) = (3c^3 - a^3)/(c^2(3c + a)).$$

The probability that 1 ends up on A and 2 on B is

$$(1 - x^*)(1 - p)^2 = a(c + a)^2/2c^2(3c + a).$$

And the probability that 1 ends up on B and 2 on A is

$$(1 - x^*)p^2 = a(c - a)^2/2c^2(3c + a).$$

We thus see:

PROPOSITION 1: V(n) is strictly increasing in n: more rounds of coordination are better. As n increases, V(n) converges to c. In early rounds of negotiation, there is almost no concession. Later on, concessions are more likely.

2.3 Bandwagon Game With n Periods

Now consider the analogous case but with unilateral adoption rather than negotiation as the coordinating principle. Just as above, we reduce things to a two-by-two game with an endogenous payoff entry, and find that we can say a good deal about the outcome from that. Let the value to each player be W(n). Then

W(n) is the value of the mixed-strategy equilibrium in the game with payoff matrix:

		Player 2	
		Wait	Commit to B
Player 1	Commit to A	a+c, c	a, a
	Wait	W(n-1), W(n-1)	a+c, c

Figure 3: Bandwagon Game with n Periods Remaining

The main difference between this game and that for the committee in Figure 2, is that it is possible for this game to terminate with incompatibility before the final period. Thus, when both players (simultaneously) commit to their own preferred technology, each gets a payoff of a.

Defining q to be the probability that a player waits, we get:

$$W(n) = q(n)(a + c) + [1 - q(n)]a = q(n)W(n - 1) + [1 - q(n)]c.$$

Again by induction, $W(n - 1) < W(n) < c$. Since

$$q(n)[2c - W(n)] = c - a,$$

$q(n)$ increases in n, towards some limit, q^* say. Let $W(n)$ converge to W^*, which cannot exceed c. Then

$$W^* = q^*(a + c) + [1 - q^*]a = q^*W^* + [1 - q^*]c,$$

whence (since W^* cannot be $a + c$)

$$W^* = c,$$

$$q^* = (c - a)/c.$$

In contrast to the committee case, then, it is quite likely that something will happen early on, even if there is

still a good deal of time remaining. While this has its good aspects, it is also dangerous: an advantage of the committee system is that incompatibility is never inevitable until the end.

As above, we can calculate the overall probabilities of various eventual outcomes. For a large enough number of periods, since q(n) converges to q^*, the outcome is close to that which would result if we simply had many periods with each side playing q^* in each period. Thus if y is the probability of coordination, we have $y = 2q^*(1 - q^*) + q^{*2}y$, whence

$$y = (2c - 2a)/(2c - a).$$

With the remaining probability (1 - y), the outcome is that player 1 adopts A and player 2 adopts B: the opposite never happens in the bandwagon system with many periods.[25] Comparing these results, we see that:

PROPOSITION 2: In a long game, the committee is always more likely to achieve coordination than the bandwagon system. It is also more likely to reach the worst outcome (1 on B, 2 on A). Hence, it is less likely that 1 chooses A and 2 chooses B.

Although the limiting payoffs for the players are the same (c) in the two systems, Proposition 2 shows that the committee is socially superior if firms have less incentive than

---

[25] We can check these calculations by noting that
$$y(2c + a) + (1-y)2a = 2a + y(2c - a) = 2c,$$
which is indeed the aggregate total payoff in a long bandwagon game.

they should to achieve compatibility.[26] Conversely, if there is too much private incentive to standardize[27], so that 2c exceeds the social benefit from compatibility, then the bandwagon system is socially superior in the limit.

We have shown that as the number of periods available for coordination increases without limit, the two methods of coordination give equal payoffs (c) in the limit. But in practice "periods" are often quite long: for committees, a period is the time that passes between meetings, often several months.[28] In the case of bandwagons, a period is the length of time between a firm's irrevocable commitment to a system and its rival's unambiguous observation of that commitment. That too may easily be months, since it is easy to see that if players were to believe verbal announcements of commitment then it would always be in a player's interest to make them, whether true or not. And not only are the periods fairly long, but there is often little time available. Therefore we examine the relative values of committees and bandwagons with finitely many periods.

To do so, we compare Figures 2 and 3 above. Call "commit" in the bandwagon game, and "insist on your preferred system" in

---

[26] As Katz and Shapiro (1985) suggest.

[27] As suggested by Farrell and Saloner (1986a), Matutes and Regibeau (1986).

[28] For decisions that must be ratified in plenary sessions, the period can be several years. For example, the plenary session of the CCIR in May 1986 failed to reach agreement on a standard for high-definition television, so that the next possible date will be in 1990. (See Besen and Johnson (1986)).

the committee game, the "tough" strategies; and call "wait" in bandwagons and "concede" in committee "soft" strategies. In each game, if (at some stage) just one player is tough, then his most-preferred outcome happens. If both sides are tough, then in the bandwagon game it is disastrous: they are committed to incompatibility; whereas in the committee game, they simply continue to the next period. Finally, if both are "soft," then in both the bandwagon game and the committee game the game continues, but with less time to reach agreement.

Thus both the committee and bandwagon games are of the form:

		Player 2	
		Soft	Tough
Player 1	Tough	a+c, c	z, z
	Soft	w, w	c, a+c

Figure 4: General Class of Coordination Games

Since V(n-1) > a, if we assume inductively that V(n - 1) > W(n - 1), then the off-diagonal elements z and w are both higher in the committee game than in the bandwagon game. If it were generally true that the equilibrium value of a game is monotonically increasing in the payoffs, we could conclude by induction that the committee is always better than the bandwagon. Unfortunately, it is easy to construct examples of games in which one can increase certain payoffs and actually reduce the equilibrium payoff, since equilibrium strategies change.

However, it is straightforward to prove the following:

PROPOSITION 3: The mixed-strategy equilibrium payoff of the game in Figure 4 is increasing in w and z if $w < c$ and $z < c$.

To see this, suppose in contradiction, that w increases but the value of the game decreases. Since $c > w$, from the bottom row this implies that Player 2 must be playing left with higher probability than before. But in that case Player 1 can do better than before by playing top (since $a+c > z$). Similarly for z.

We therefore conclude:

PROPOSITION 4: With a finite fixed horizon, the committee unambiguously outperforms the bandwagon system.

Recall that the above analysis assumed that if both players concede in the committee game then they have failed to coordinate, and must reconvene in the committee a period later. As we mentioned above, this seems unduly harsh. If both players show a willingness to concede it seems more reasonable that some coordinated action be achieved. In that case w in Figure 4 is even higher in the committee game so that, by Proposition 3, the committee does even better.

3. Allowing for the Value of Speed

In Section 2 we assumed that players care only about what eventually happens, not about when. In some cases, this assumption is reasonable, but often time is valuable and players would rather reach agreement sooner than later. Since the committee is slower, one might expect that this could reverse the result of the previous section. In this section, therefore, we allow for such impatience. Specifically, we suppose that

payoffs are discounted by a factor d < 1 for each period that decisions are postponed.

In the committee system, each period (until agreement is reached) the players face a game of the form:

		Player 2	
		Concede	Insist
Player 1	Insist	a+c, c	dV, dV
	Concede	dV, dV	c, a+c

Figure 5: Infinite Horizon Committee Game with Discounting

where V is now the value of being at the beginning of a period without agreement.[29] Similarly, in the bandwagon system, the period-game has the form:

		Player 2	
		Wait	Commit
Player 1	Commit	a+c, c	a, a
	Wait	dW, dW	c, a+c

Figure 6: Infinite Horizon Bandwagon Game with Discounting

where W is the value of being in that game. Since we now have a stationary system, we cannot work by induction as in Section 2.

---

[29] As before, the results only become stronger if we make a more generous assumption about what happens if both concede.

Instead, to compare these equilibria, consider the class of games:

		Player 2	
		Soft	Tough
Player 1		a+c, c	t, t
		dU, dU	c, a+c

Figure 7: General Infinite Horizon Game

where U(t) is the mixed-strategy equilibrium value of the game. Let r(t) denote the probability that player 2 plays soft. Then

$$U(t) = r(t)(a + c) + [1-r(t)]t \qquad (2a)$$

$$= r(t)dU(t) + [1-r(t)]c. \qquad (2b)$$

Rearranging (2b) and differentiating gives

$$U'(t) = cr'(t)(d-1)/[1-r(t)d]^2.$$

Since $d < 1$, this means that $U'(t)$ and $r'(t)$ have opposite signs. Differentiating (2a),

$$U'(t) = r'(t)(a + c - t) + [1 - r(t)].$$

If $a + c > t$, for $U'(t)$ and $r'(t)$ to have opposite signs it must be that:

PROPOSITION 5: U(t) is increasing, and r(t) is decreasing, in t.

Now, it must surely be the case that $dV > a$ or else the players would never agree to play the committee game at all. Therefore the "t" in the committee period-game exceeds that in the bandwagon period-game which implies:

PROPOSITION 6: The committee system is unambiguously better than the bandwagon system when there is no fixed deadline but delay is costly.

Since r(t) is decreasing in t, "tough" behavior is more common in the committee system. However, it is less harmful.

This is, of course, very much like the behavior of the two systems in the analysis with a fixed deadline but many periods. In fact the infinite horizon problem with d = 1 and the finite horizon problem with n infinite are identical, as can be seen by comparing the period-by-period payoff matrices. We thus see that the two systems perform equally well (though descriptively differently) when time is of no concern; but the committee system is better when time is important.

This may at first seem somewhat surprising, given that the committee is slower. One might think that the slower, more reliable system would seem less attractive when we put in a value for time. However, what counts, of course, is how the equilibrium payoffs change as we lower d from 1. An immediate effect of doing this is that the committee players no longer play tough with probability one. As a result, the committee acts much more quickly than before. Bandwagon players also become less patient, of course, and so resolve matters more quickly as well. However, when both bandwagon players are playing tough with higher probability, this gain in speed must be offset against the increased cost of disastrous uncoordinated outcomes.

4. Hybrid Coordination

The committee system as analyzed above demands a commitment from participating agents not to adopt their proposed solutions unilaterally, before the committee reaches a decision. Otherwise, if other players are mired in talk, one agent can simply commit, and thus obtain his most preferred outcome, since it is common knowledge that the others then will follow his bandwagon.

Often, however, such commitment is not available, not credible, or not undertaken. It is commonplace for committee deliberations to be spurred, threatened, disrupted and reversed by marketplace developments in which a participant refuses to await a cooperative decision any longer, and begins manufacturing and selling its system. Thus the "pure" committee system which we have analyzed is an abstraction.

The pure bandwagon game is an abstraction also. It is unlikely that agents in a market with large compatibility benefits will determine in advance that they will not communicate. So, although we find it helpful to analyze the pure cases, it is also important to analyze what happens if both mechanisms coexist. That is the purpose of this section.

We model the hybrid game as follows: In each period, each player has three choices. He can irrevocably adopt his preferred solution, hoping that the other will then follow. Or, if he refrains from that, he can go to committee where he has his usual options of "insist" and "concede".

It turns out that the game in which the choice of these three actions are made simultaneously by both players, and that in which they first decided whether to adopt and then, if they go to committee whether to insist or concede, are equivalent because adoption is an irrevocable commitment. We use the two stage formulation since it is better for expository purposes.

Using the infinite horizon formulation with discounting introduced in the previous section, we can represent the two stages of the hybrid game as follows:

First stage:

		Player 2	
		Committee	Adopt
Player 1	Adopt	a+c, c	a, a
	Committee	H', H'	c, a+c

Second Stage:

		Player 2	
		Concede	Insist
Player 1	Insist	a+c, c	dH, dH
	Concede	dH, dH	c, a+c

Figure 8: The Hybrid Game

Here, H is the value of the hybrid game at the beginning of a period, while H' is the value of the hybrid game in any period after both players have decided not to adopt in that period.

Note that the first stage has the typical structure of the bandwagon game, while the second stage has that of the committee game. One might be tempted to conclude that the payoffs are also

a hybrid of those found in the pure games. Suprisingly, the payoffs in the hybrid game are the highest of all! To see this we consider games of the form:

		Player 2	
Player 1		a+c, c	dt, dt
		dt, dt	c, a+c

Figure 9: Symmetric Infinite Horizon Game

Let the value of this game be X(t) and let the probability with which Player 2 plays "left" be f(t). Simple calculations show that $X(0)=(a + c)c/(a + 2c) > 0$. Also

$$X(c) = f(t)dc + [1 - f(t)]c < c.$$

Finally, we can show that $0 < X'(t) < 1$: Observe first that $f(t)(a + c) + (1 - f(t))dt = f(t)dt + (1 - f(t))c$, and so

$$f(t) = (c - dt)/(a + 2(c - dt)) < 1/2.$$

Therefore, since $X(t) = f(t)dt + [1 - f(t)]c$,

$$X'(t) = df(t) + f'(t)(dt - c),$$

which after some calculation,

$$= df(t)[1 + a/(a + 2(c - dt))].$$

Now, since $d < 1$ and $f(t) < 1/2$, $X'(t) < 1$. Also, by inspection, $X'(t) > 0$. Therefore we have:

PROPOSITION 7: If X(t) is the value of the game in Figure 9, then $X(0) > 0$, $X(c) < c$, and $0 < X'(t) < 1$.

Therefore, X(t) is as illustrated in Figure 10. X(t) has a unique fixed point. By inspection of Figure 5, it is the value of the committee game, V.

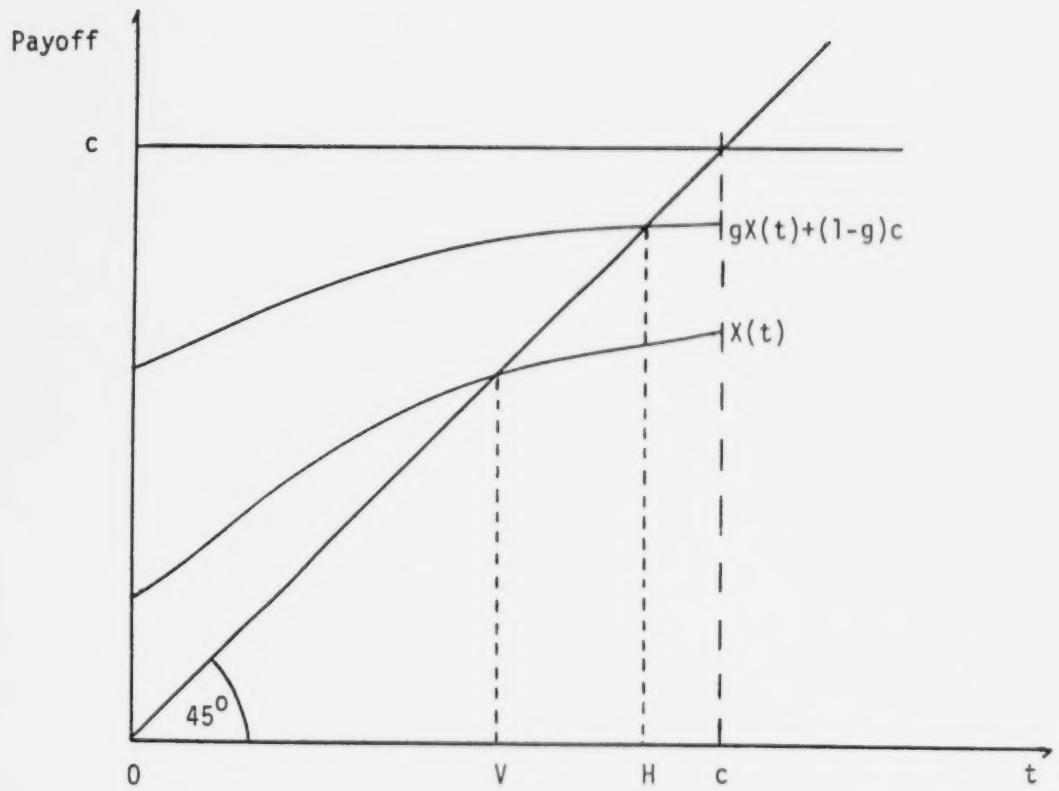

Figure 10: Committee and Hybrid Games Compared

Now consider the hybrid game in Figure 8. From the first stage, we have $H = gH' + (1 - g)c$, where g is the probability the players play "committee". Comparing the second stage with the game in Figure 9, we see that $H' = X(H)$. Combining these expressions gives $H = gX(H) + (1 - g)c$. Therefore H is the solution of the expression $t = gX(t) + (1 - g)c$. Since, as we have argued above, g is positive, and since $V < c$, the right-hand-side of this equation is plotted on Figure 10. From this it is apparent that $H > V$; the hybrid system outperforms the committee.

A heuristic intuition for this result is that the hybrid system gives two chances for coordination each period rather than just one. We can evaluate player 1's payoff by assuming that he plays committee in the first stage. His expected payoff, H, is thus a mixture of c (which he gets if player 2 adopts) and H' (which he gets if player 2 also chooses committee). The pure committee game and the second stage of the hybrid game differ only in that dH replaces dV in the continuation cells. If $H > V$, then the second stage of the hybrid game has a higher payoff than the pure committee game, i.e., $H' > V$. In that case, H is a mixture of two payoffs, both of which exceed V, and so itself exceeds V. This suggests that an equilibrium can have $H > V$ and that it might be hard to have $H < V$. This intuition is confirmed by the proof above.

A different way to see the intuition for the result is the following: Suppose that at the first stage both players were

going to choose committee with probability one, but decide instead to each play "adopt" with a very small probability, e. In that case they face the risk that they will not achieve a coordinated outcome. However, the probability that this will occur is only e^2 - a second-order effect. On the other hand, with probabilty 2e(1-e) - a first-order effect - only one of them adopts, giving a strictly higher payoff. So if, as it turns out, the equilibrium value of e (i.e., g) is not too large, then they are better off.

5. Conclusion

Problems of coordination arise in a variety of important economic and other contexts. For example, coordination among potential entrants to an industry is an important assumption of the usual theory of entry, aggregate demand and employment decisions involve coordination by firms and workers[30], the theory of economic development emphasises problems of coordination in the development of different sectors,[31] and efficiency in monopolistic competition can require coordination by manufacturers of complementary products[32].

Active coordination can take at least four forms. The first is tradition, if tradition can consistently point to one outcome.

---

[30] See Diamond (1984) and Roberts (1986) for examples in which coordination difficulties can cause aggreagate unemployment.

[31] See for instance Rosenstein-Rodan (1943).

[32] See Hart (1980).

For instance, certain kinds of animals[33] always return to a fixed place at a fixed time of year for mating. But this method breaks down when the problems faced are new, in which case tradition gives inadequate guidance.

The second form of coordination is authority: if one player is nominated in advance to choose, then others can defer to him. However, there may be no authority figure in a given interaction, or there may be more than one. We can sometimes nominate "government" as a universal authority figure, but even that breaks down when problems are international.

Then there are the two methods of coordination studied in this paper. One, the bandwagon system, involves waiting for one player to move first, so that the other can follow and coordination be achieved that way. If interests conflict, however, the first mover has extra power[34], and in equilibrium there is always a dissipative fight over who goes first. In our model, that took the form of excessive haste to move (even when there is plenty of time) so that unnecessary coordination failures result. The other, the committee system, involves talking. Despite being extremely widespread, this method has been neglected in the economics literature.[35]

---

[33] Elephant seals, for example.

[34] See the "New Hampshire Theorem" (Proposition 2) in Farrell and Saloner (1985).

[35] Farrell and Saloner (1985), Farrell and Gibbons (1986), and Farrell (1987) are the only explicit applications of this idea in economics that we know of. Lewis (1969) bases an entire theory of language on the need for coordination of action.

These latter two methods of achieving coordination are particularly relevant where neither tradition nor an authority structure creates a decisive asymmetry. We have studied a simple model of how these two schemes of coordination work, and have compared them descriptively and normatively. Finally, we have evaluated a hybrid system in which both actions and words are used, and found it more effective than either pure system.

Standardization for compatibility is an important activity, especially in information-processing industries. Previous literature focused exclusively on methods of standardization that do not involve explicit coordination. This paper is a beginning of a more balanced and accurate theory of standardization.

REFERENCES

Besen, S.M. and L.L. Johnson, "Compatibility Standards, Competition, and Innovation in the Broadcasting Industry," The Rand Corporation, R-3453-NSF, November 1986.

Besen, S.M. and G. Saloner, " Compatibility Standards and the Market for Telecommunications Services," Working Papers in Economics E-87-15, The Hoover Institution, March 1987.

Carlton, D.W. and J.M. Klamer, "The Need for Coordination Among Firms, With Special Reference to Network Industries," The University of Chicago Law Review 50 (1983), 446-465.

Crane, R., The Politics of International Standards, Norwood, N.J.: Ablex, 1979.

Diamond, P., A Search-Equilibrium Approach to the Foundations of Macroeconomics, Cambridge, MA: M.I.T. Press, 1984.

Farrell, J. "Cheap Talk, Coordination, and Entry," Rand Journal of Economics 18 (Spring 1987), 34-39.

Farrell, J. and R. Gibbons, "Cheap Talk in Bargaining," M.I.T. Working Paper, 1986.

Farrell, J. and G. Saloner, "Standardization, Compatibility, and Innovation," Rand Journal of Economics 16 (1985), 70-83.

_____, "Standardization and Variety," Economics Letters 20 (1986a), 71-74.

_____, "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation", American Economic Review 76 (1986b), 940-955.

Fudenberg, D. and J. Tirole, "Preemption and Rent Equalization in the Adoption of New Technology," Review of Economic Studies 52 (1985), 383-402.

Hart, O., "Perfect Competition and Optimal Product Differentiation," Journal of Economic Theory 22 (1980), 279-312.

Hendricks, K. and C. Wilson, "Equilibrium in Preemption Games with Complete Information," Working Papers in Economics E-86-72, oover Institution, 1986.

Katz, M.L. and C. Shapiro, "Network Externalities, Competition, and Compatibility," American Economic Review 75 (1985), 424-440.

_____, "Technology Adoption in the Presence

of Network Externalities," Journal of Political Economy 94 (1986a), 822-41.

_____, "Product Compatibility Choice in a Market with Technological Progress," Oxford Economic Papers 38 (1986), 146-165.

Lewis, D., Convention, Cambridge, MA: Harvard University Press, 1969.

Matutes, C. and P. Regibeau, " 'Mix and Match': Product Compatibility Without Network Externalities," UC Berkeley, May 1986, mimeo.

Rosenstein-Rodan, P., "Problems of Industrialization of Eastern and Southeastern Europe," Economic Journal (1943).

Rutkowski, A.M., Integrated Services Digital Networks, Dedham, MA: Artech House, 1985.

_____, "An Overview of the Forums for Standards and Regulations for Digital Networks," Telecommunication (North American Edition) 20 (1986), 84-96.

Sanders, T.R.B., The Aims and Principles of Standardization, International Organization for Standardization, 1972.

Sullivan, M.T., R.T. Zader, "The Role of Standards in Network Evolution," Computers and Standards 4 (1985), 33-53.

Verman, L.C., Standardization, A New Discipline, Hamden: Ardon Books, 1973.